AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD023889

Key Points:
- To assess basin-scale energy components in NCEP operational and research NLDAS-2
- To jointly use GEWEX/SRB radiation and gridded FLUXNET-based energy fluxes
- To directly evaluate NLDAS-2 energy products at 12 CONUS River Forecast Centers

Supporting Information:
- Text S1

Correspondence to:
Y. Xia,
youlong.xia@noaa.gov

Citation:
Xia, Y., B. A. Cosgrove, K. E. Mitchell, C. D. Peters-Lidard, M. B. Ek, S. Kumar, D. Mocko, and H. Wei (2016), Basin-scale assessment of the land surface energy budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems, *J. Geophys. Res. Atmos.*, *121*, 196–220, doi:10.1002/2015JD023889.

Received 2 JUL 2015
Accepted 10 DEC 2015
Accepted article online 15 DEC 2015
Published online 14 JAN 2016

Basin-scale assessment of the land surface energy budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems

Youlong Xia[1,2], Brian A. Cosgrove[3], Kenneth E. Mitchell[4], Christa D. Peters-Lidard[5], Michael B. Ek[1], Sujay Kumar[5,6], David Mocko[5,6], and Helin Wei[1,2]

[1]Environmental Modeling Center (EMC), National Centers for Environmental Prediction (NCEP), College Park, Maryland, USA, [2]I. M. System Group (IMSG) at NCEP/EMC, College Park, Maryland, USA, [3]National Water Center, National Weather Service (NWS), Silver Spring, Maryland, USA, [4]Prescient Weather Ltd., State College, Pennsylvania, USA, [5]Hydrological Sciences Laboratory, NASA Goddard Space Flight Center (GSFC), Greenbelt, Maryland, USA, [6]SAIC at NASA/GSFC, Greenbelt, Maryland, USA

Abstract This paper compares the annual and monthly components of the simulated energy budget from the North American Land Data Assimilation System phase 2 (NLDAS-2) with reference products over the domains of the 12 River Forecast Centers (RFCs) of the continental United States (CONUS). The simulations are calculated from both operational and research versions of NLDAS-2. The reference radiation components are obtained from the National Aeronautics and Space Administration Surface Radiation Budget product. The reference sensible and latent heat fluxes are obtained from a multitree ensemble method applied to gridded FLUXNET data from the Max Planck Institute, Germany. As these references are obtained from different data sources, they cannot fully close the energy budget, although the range of closure error is less than 15% for mean annual results. The analysis here demonstrates the usefulness of basin-scale surface energy budget analysis for evaluating model skill and deficiencies. The operational (i.e., Noah, Mosaic, and VIC) and research (i.e., Noah-I and VIC4.0.5) NLDAS-2 land surface models exhibit similarities and differences in depicting basin-averaged energy components. For example, the energy components of the five models have similar seasonal cycles, but with different magnitudes. Generally, Noah and VIC overestimate (underestimate) sensible (latent) heat flux over several RFCs of the eastern CONUS. In contrast, Mosaic underestimates (overestimates) sensible (latent) heat flux over almost all 12 RFCs. The research Noah-I and VIC4.0.5 versions show moderate-to-large improvements (basin and model dependent) relative to their operational versions, which indicates likely pathways for future improvements in the operational NLDAS-2 system.

1. Introduction

The North American Land Data Assimilation System (NLDAS) is a multi-institutional collaborative project [*Mitchell et al.*, 2004] that has completed two major phases so far. Phase 1 (NLDAS-1) was initiated in 1999 [*Mitchell et al.*, 1999] with the purpose of establishing a land data assimilation framework for the continental U.S. (CONUS) that includes several land surface models (LSMs), high-quality meteorological forcing, soil and vegetation parameters, and validation data sets. The NLDAS-1 simulations were conducted from 1 October 1996 to 30 September 1999 at a 0.125° resolution over the NLDAS domain, which encompasses the CONUS, the southern part of Canada, and the northern part of Mexico (25–53°N). Details on NLDAS-1 are documented in *Mitchell et al.* [2004]. Phase 2 (NLDAS-2) is an extension [*Xia et al.*, 2012] of the NLDAS-1 with the key goals of producing quality-controlled and reliable long-term (>30 years) and real-time hydrometeorological products, including water and energy budget components to support the National Integrated Drought Information System (drought.gov), national operational drought monitoring and prediction tasks, and water resource information needs of various government agencies, academia, and private enterprise. To provide timely and stable products to all users, NLDAS-2 became part of the National Centers for Environmental Prediction (NCEP) operational suite in August 2014 (initial operational NLDAS-2 version). A research version of NLDAS-2 continues to be maintained by the Environmental Modeling Center (EMC) Land-Hydrology group. The research version focuses on incorporating improvements to the LSMs and their common forcing fields, plus adding data assimilation capabilities, from both EMC in-house efforts and efforts of our collaborators and partners, with the

Table 1. Primary Attributes of the Three of Four Land Models in NLDAS-2 That Provide Surface Energy Budget Components, Modified From *Mitchell et al.* [2004][a]

Description	Mosaic [*Koster and Suarez*, 1996]	Noah [*Ek et al.*, 2003]	VIC [*Liang et al.*, 1994]
		Operational NLDAS-2 Models	
Energy balance	Yes	Yes	Yes
Model time step	15 min	15 min	1 h
No. of model soil layers	3	4	3
Model soil layer thicknesses	10, 30, and 160 cm	10, 30, 60, and 100 cm	10 cm, variable
Tiling: vegetation	Yes	No	Yes
Tiling: elevation	No	No	Yes
Soil thermodynamics	Partial force restore	Heat conduction equation	Heat conduction equation modified
Soil temperature profile	No	Yes	Yes
Snow-free albedo	Vary with LAI, GVF, and biome	Monthly input background field	Vary with LAI and biome
Diurnal albedo	Yes	No	No
Explicit vegetation	Yes	Yes	Yes
Canopy resistance	*Sellers et al.* [1986]	*Jarvis* [1976]	*Jarvis* [1976]
Root depth	40 cm	Variable (1 or 2 m)	Variable (1.35–3 m)
Note	The version is the same as used in NLDAS-1.	Snow model and warm season model physics upgrade [*Livneh et al.*, 2010; *Wei et al.*, 2013].	The version is the same as used in NLDAS-1.
		Research NLDAS-2 Models	
Description	Mosaic	Noah-I	VIC4.0.5
Upgraded	No	Yes. Model was modified by *Xia et al.* [2014c].	Yes. Soil parameters were calibrated in *Troy et al.* [2008].

[a]LAI is leaf area index, and GVF is greenness vegetation fraction.

goal of transitioning the advancements into the next phases of the operational NLDAS. In this study, the research version includes only LSM improvements, relative to the operational NLDAS. These LSM improvements are detailed in the NLDAS-2 model section (section 2).

One of the key efforts of the NLDAS-2 project is to evaluate the ability of LSMs to simulate the energy and water budgets on river basin scales by carrying out validation using both in situ and satellite-based observations. LSMs have many physical cycles (e.g., water, energy, carbon, and nitrogen) with the water and energy cycles viewed as the fundamental cycles that underpin the others. The water cycling processes [*Getirana et al.*, 2014] describe the partitioning of precipitation into evapotranspiration, total runoff, and change in total water storage (i.e., changes in snow water equivalent, canopy water storage, and total soil water storage). The energy cycling processes describe the partitioning of the surface net radiation into sensible heat flux, latent heat flux, and downward ground heat flux to the subsurface via conduction. The water and energy cycling processes are obviously intertwined because the evapotranspiration partitioned from precipitation is constrained by the energy budget equation. In addition, changes in soil moisture, soil temperature [*Xia et al.*, 2013], and land surface skin temperature also modulate water and energy cycling. By upgrading the surface water and/or energy budget components of the LSMs, improvements can be made in the retrospective and real-time NLDAS-2 products. The evaluation of the surface water components in both the operational and research NLDAS-2 over large river basin scales CONUS wide will be reported in *Xia et al.* [2015a]. Evaluations of the surface energy budget components at limited spatial and temporal scales have been performed earlier [*Xia et al.*, 2012, 2014a, 2014b], but we emphasize that the latter studies focused on in situ radiation flux observations of the Southern Great Plains region and investigated the monthly mean diurnal energy cycle for only a relatively short period (1997–1999). Subsequently, a more comprehensive evaluation of the surface energy budget terms in either the operational or research NLDAS-2 remained a challenge due to the lack of long-term (e.g., $>$20 years) observational information, in particular at large river basin scales (e.g., National Weather Service's River Forecast Centers). Thus, the primary purpose of the new surface energy budget study here (and the main difference from its forerunner studies) is that for the first time the focus is on the *long-term* (multidecade) monthly radiation fluxes at large river basin scales CONUS wide. Important additional purposes of this study are (1) to further promote, via the validation results, the application of the NLDAS-2 products

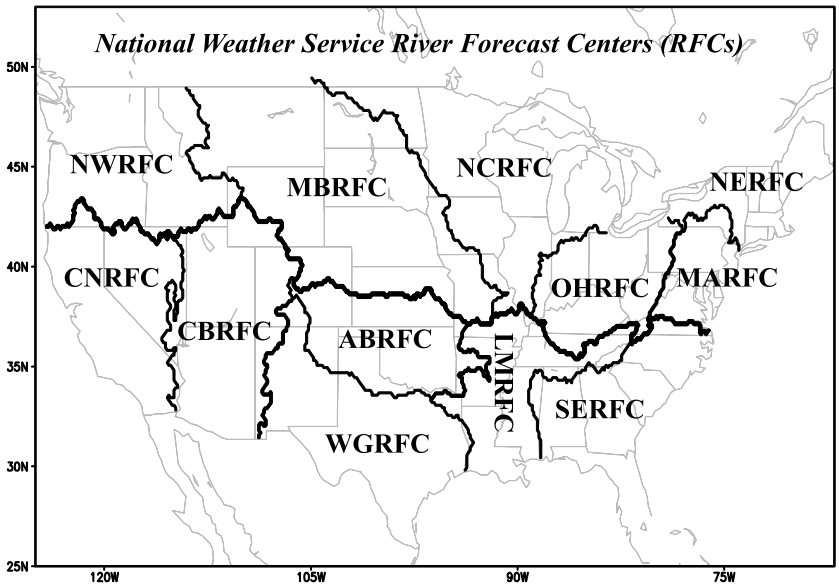

Figure 1. Boundaries of the domains of the 12 CONUS National Weather Service River Forecast Centers (RFCs).

nationally by the user community; (2) to justify the use of the newer versions of the LSMs demonstrated here; and (3) to suggest future pathways in developing the next phase of NLDAS.

The role of surface energy budgets is critical in determining the boundary conditions that control weather and climate at different time scales [*Berbery et al*., 1999]. Downward shortwave and longwave radiation determine how much input energy is available at the land surface. Upward shortwave radiation is determined by the surface albedo and the incoming shortwave radiation, whereas upward longwave radiation depends on the land surface temperature and emissivity. The difference of net shortwave and longwave radiation—the net radiation—is the primary energy input to the land surface, and it directly affects the land surface skin temperature and heat fluxes. The imbalances between incoming and outgoing radiation serve as a primary control on evapotranspiration and therefore the hydrological cycle [*Wild et al*., 2008]. Understanding how well we can simulate the partitioning of net radiation, in particular in terms of mean annual cycles and seasonal variation, is a key objective of NLDAS-like studies because the net radiation controls the seasonal variation of sensible heat flux, latent heat flux, and ground heat flux. Therefore, many articles have evaluated the performance of uncoupled and coupled LSMs in simulating various surface energy budget processes at river basin, regional and global scales [*Betts et al*., 1998, 1999; *Berbery et al*., 1999; *Roads et al*., 2003; *Feng and Houser*, 2008; *Boone et al*., 2009; *Troy and Wood*, 2009]. However, a comprehensive evaluation of radiation, sensible heat, and latent heat fluxes for uncoupled LSMs executed on a distributed regional/global domain over two to three decades has not been performed yet due to lack of appropriate reference data sets that span both the long temporal period (decades) and large spatial domain (CONUS wide).

Table 2. Names of the 12 CONUS RFCs and Their Corresponding Mean Annual, Basin Average Values of Bowen Ratio (SH/LH), Calculated Over the 24.5 year Period (July 1983 to December 2007) From the FLUXNET-Based MTE Database for SH and LH, and 30 year (1960–1990) Ratio of Precipitation (P) to Potential Evapotranspiration (PE)[a]

RFC Label	RFC Name	Bowen Ratio	P/PE
CBRFC	Colorado	2.91	0.29
CNRFC	California-Nevada	2.19	0.37
WGRFC	West Gulf	1.31	0.37
MBRFC	Missouri	0.98	0.50
ABRFC	Arkansas	0.97	0.54
NCRFC	North central	0.59	0.82
NWRFC	Northwest	1.30	0.96
MARFC	Mid-Atlantic	0.67	1.03
SERFC	Southeast	0.52	1.04
NERFC	Northeast	0.81	1.22
LMRFC	Lower Mississippi	0.53	1.29
OHRFC	Ohio	0.60	1.33

[a]The P/PE values are adapted from *Schaake et al*. [2004]. The RFCs are listed in increasing order of P/PE.

Table 3. Reference Values of Mean Annual Surface Energy Fluxes for the 12 CONUS RFCs Over the 24.5 year Period (July 1983 to December 2007)[a]

RFC Label	DSWR ($W\,m^{-2}$)	USWR ($W\,m^{-2}$)	DLWR ($W\,m^{-2}$)	ULWR ($W\,m^{-2}$)	R_n ($W\,m^{-2}$)	SH* ($W\,m^{-2}$)	LH* ($W\,m^{-2}$)	$\frac{SH+LH}{R_n}$ [−]
CBRFC	208.5	27.3	289.6	386.7	74.0	63.1 ± 15.8	21.7 ± 6.1	1.14
CNRFC	209.6	31.3	297.3	394.2	81.4	59.0 ± 14.0	26.9 ± 6.6	1.06
WGRFC	205.8	31.8	333.3	410.7	96.7	54.2 ± 12.5	41.4 ± 9.2	0.99
MBRFC	171.0	28.8	295.5	357.5	80.2	34.1 ± 6.2	34.8 ± 6.4	0.86
ABRFC	194.5	28.2	320.6	389.5	97.4	42.5 ± 8.9	42.9 ± 8.3	0.88
NCRFC	156.6	23.3	303.1	352.2	84.2	25.2 ± 5.5	42.8 ± 5.1	0.81
NWRFC	170.5	24.9	285.3	353.2	77.7	41.7 ± 7.8	32.1 ± 6.1	0.95
MARFC	158.5	19.1	320.6	370.5	89.6	33.4 ± 6.1	50.0 ± 6.4	0.93
SERFC	186.1	20.6	355.0	409.4	111.0	33.4 ± 7.1	64.5 ± 6.7	0.88
NERFC	148.9	18.9	301.8	346.3	85.5	33.1 ± 6.3	41.0 ± 6.5	0.87
LMRFC	179.6	20.7	348.4	402.4	104.9	32.9 ± 7.3	62.0 ± 8.4	0.91
OHRFC	160.9	21.1	322.9	375.1	87.6	30.6 ± 6.2	51.1 ± 6.1	0.93

[a]The first five columns give downward shortwave radiation (DSWR), upward shortwave radiation (USWR), downward longwave radiation (DLWR), upward longwave radiation (ULWR), and net radiation (R_n) calculated from SRB/GEWEX satellite observations. The last three columns give sensible heat flux (SH), latent heat flux (LH), and the estimate of the energy budget closure error calculated from the FLUXNET-based MTE database. The closure error is calculated as the ratio $[(SH + LH)/R_n]$ assuming that the mean annual ground heat flux and heat storage in the vegetation canopy are negligible. (In support of this assumption, for each of the 12 RFCs the corresponding mean annual NLDAS-2-simulated ground heat flux of the Mosaic and Noah LSMs is within 1.0 $W\,m^{-2}$ and for the VIC LSM is within 3.0 $W\,m^{-2}$).

Recently, the NASA Langley Research Center released the 24.5 year (July 1983 to December 2007) GEWEX/SRB (Global Energy and Water Cycle Experiment/Surface Radiation Budget) monthly mean surface shortwave and longwave flux data sets on a regular 1.0° global grid (http://eosweb.larc.nasa.gov, version 3.1). These radiation fluxes were validated by *Zhang et al*. [2013]. The Max Planck Institute (MPI) in Germany has released monthly sensible and latent heat fluxes on a regular 0.5° global grid [*Jung et al*., 2009, 2011] generated by using a multitree ensemble method (MTE) and FLUXNET tower flux observations [*Baldocchi et al*., 2001]. These newly released data sets provide an opportunity to perform a comprehensive evaluation for both the operational and research NLDAS-2.

We recognize that there is a spatial-scale conflict among NLDAS-2 energy budget components, GEWEX/SRB radiation fluxes, and MTE sensible and latent heat flux products. To minimize the uncertainties and errors caused by spatial-scale conflicts, we spatially average these different resolution products over the 12 National Weather Service (NWS) River Forecast Centers (RFCs) within the CONUS. These 12 RFCs were previously used to perform relevant comparisons of the LSMs in both NLDAS-1 and NLDAS-2 [*Schaake et al*., 2004; *Sheffield et al*., 2003; *Berbery et al*., 1999; *Xia et al*., 2015a].

This paper is organized as follows. Section 2 presents descriptions of the NLDAS-2 LSMs, the NLDAS-2 products, and the independent reference data sets. Evaluations of the multimodel NLDAS-2 surface energy budget components against independent reference data sets are presented in section 3. The summary and main conclusions are provided in section 4.

2. NLDAS Models, Products, and Reference Data Sets

2.1. NLDAS-1 Versus NLDAS-2 Models

The NLDAS-1 initiated in 1999 [*Mitchell et al*., 1999] established the infrastructure, initial selection of land surface models (LSMs), generation of the common surface forcing data, collection of soil and vegetation data sets, and in situ and satellite-retrieved observations initially used in NLDAS-1. The model outputs were evaluated and compared with in situ observations and satellite-retrieved products. The overall results showed that all four LSMs of NLDAS-1 were able to capture the broad features for the validated variables such as energy fluxes (e.g., net radiation, sensible heat, latent heat, and ground heat), water fluxes (i.e., evapotranspiration and total runoff), and state variables (i.e., soil temperature, soil moisture, land surface temperature, snow cover fraction, and snow water equivalent). The validation tools and overall results are detailed in *Mitchell et al*. [2004].

Based on these findings from NLDAS-1, the follow-on NLDAS-2 focused on improving model physics, tuning model parameters, improving surface forcing data quality and reliability, and extending the short term (i.e., 3 years) of the NLDAS model products to long term (>36 years) and real time. The NCEP NLDAS team improved Noah simulations during the cold season [*Livneh et al*., 2010] and warm season [*Wei et al*., 2013]

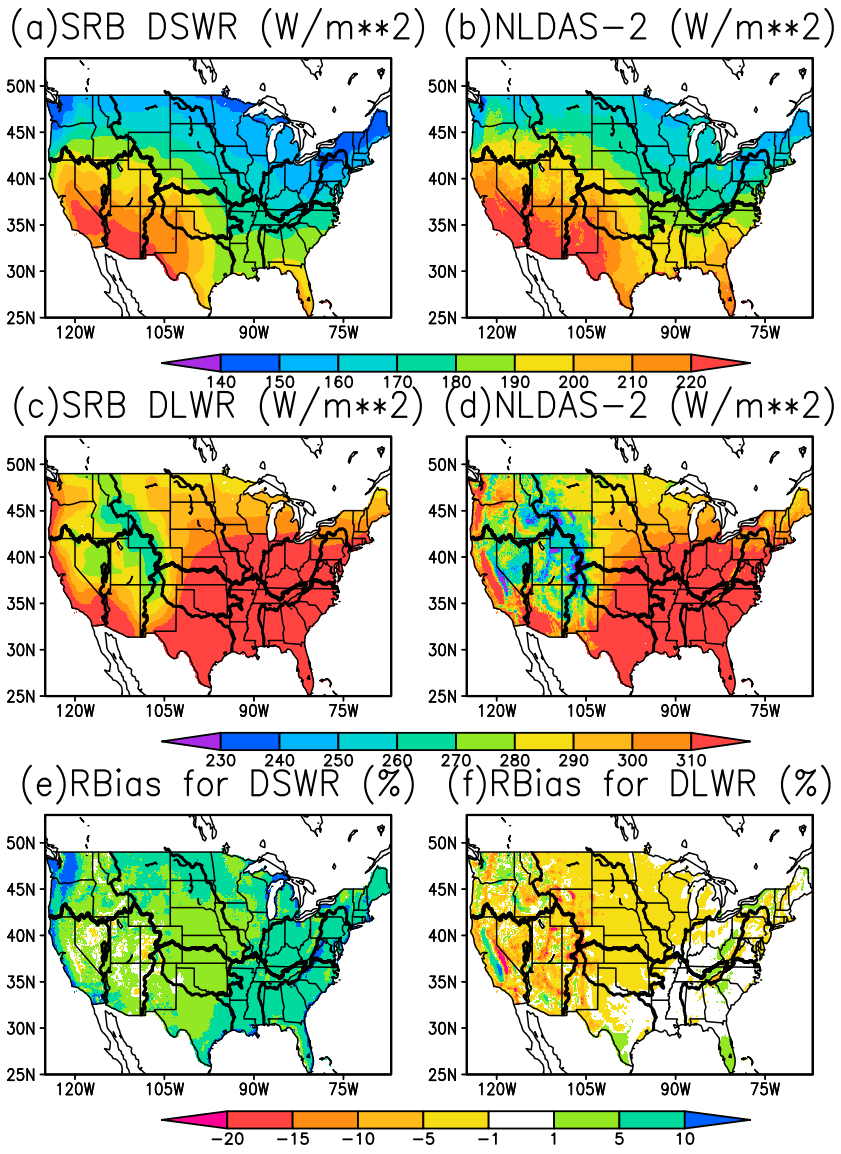

Figure 2. Spatial distribution of mean annual downward shortwave radiation (DSWR) and downward longwave radiation (DLWR) from SRB and NLDAS-2 (unit: W m^{-2}), plus the corresponding NLDAS-2 relative bias [RBias = 100*(NLDAS-2 − SRB)/SRB] for the 24.5 year period of July 1983 to December 2007. The NLDAS-2 downward radiation forcing is the same in both the operational and research NLDAS-2 systems.

through collaboration with the University of Washington. The Princeton land group improved the VIC simulation by calibrating model parameters [*Troy et al.*, 2008], and the NCEP NLDAS team also improved SAC simulations by using climatologically averaged observed potential evaporation [*Xia et al.*, 2012]. These improved LSMs were either implemented into the NLDAS operational version (i.e., Noah 2.8) or research version (i.e., SAC-Clim and VIC 4.0.5).

NLDAS-1 and NLDAS-2 include four LSMs (i.e., Mosaic, Noah, SAC, and VIC) from various communities and universities. Mosaic and Noah grew from the legacy of the surface-vegetation-atmosphere transfer (SVAT) community. SAC and VIC were developed from the hydrology community as uncoupled hydrological models with considerable calibration [*Burnash et al.*, 1973; *Liang et al.*, 1994]. The Mosaic model was developed for use in a NASA global climate model [*Koster and Suarez*, 1994, 1996]. The Noah model was developed in NCEP/EMC for application in NCEP operational regional and global numerical weather and seasonal climate forecast models [*Ek et al.*, 2003; *Meng et al.*, 2012; *Zheng et al.*, 2012; *Saha et al.*, 2014].

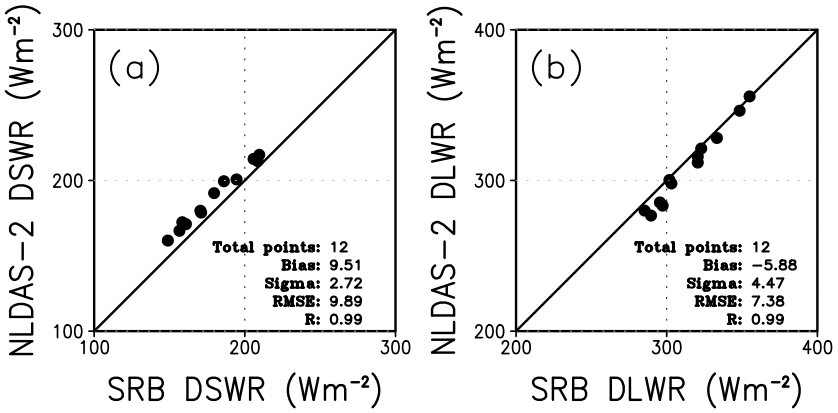

Figure 3. Scatterplots and statistical metrics for mean annual (a) downward shortwave radiation (DSWR) and (b) downward longwave radiation (DLWR) from SRB derived versus NLDAS-2 for the 12 CONUS RFCs for the 24.5 year period of July 1983 to December 2007.

SAC was developed as a lumped conceptual hydrology model [*Burnash et al*., 1973] with water cycling only and is used operationally at NWS RFCs. It was first executed as a semidistributed (grid-based) hydrology model in NLDAS-1 [*Mitchell et al*., 2004] and recently upgraded to include soil temperature calculations and surface energy budget cycling (SAC-HTET) [*Koren et al*., 2007, 2010, 2014]. While this latter version of SAC will be included in subsequent NLDAS phases (i.e., NLDAS-3), NLDAS-2 applies the nonenergy budget version of SAC.

The VIC model was jointly developed at the University of Washington and Princeton University as a macroscale semidistributed model [*Liang et al*., 1994; *Wood et al*., 1997]. VIC features both a water and energy mode and may be executed in water mode only. The water mode includes water balance cycling without an energy balance, which results in VIC behaving as a hydrological model. The VIC energy mode includes both water and energy balance cycling calculations, which results in VIC behaving as a SVAT-like (Soil-Vegetation-Atmosphere Transfer) model. In NLDAS-1 and NLDAS-2, the VIC energy mode was used to provide both water and energy cycling simulations. The four NLDAS-2 LSMs are further detailed in *Xia et al*. [2012, 2015a].

The version of VIC running in the operational NLDAS-2 is VIC4.0.3, while the research version of VIC in this study is VIC4.0.5. The latter version of VIC contains updates to its snow albedo curves and other small changes/fixes compared to VIC4.0.3. Additionally, VIC4.0.5 uses recalibrated soil parameters (relative to VIC4.0.3) developed by *Troy et al*. [2008]. The research version of Noah (labeled Noah-I in this study, for Noah "Interim") contains several updates, including changes to the treatment of the surface exchange coefficient. The surface energy fluxes simulated by VIC4.0.5 have not been evaluated against observational energy fluxes heretofore but have been compared previously with those simulated by the Mosaic and Noah LSMs [*Xia et al*., 2012]. In this study, only the three NLDAS-2 LSMs with energy cycling (Mosaic, Noah, and VIC) are included, as the version of the SAC model used in NLDAS-2 does not simulate energy components. Table 1 compares the attributes of these three operational LSMs and the two of the three LSMs with upgraded versions used in the NLDAS-2 study here.

2.2. NLDAS-2 Meteorological Forcing

NLDAS-2 hourly surface forcing fields of longwave radiation, 2 m air temperature, 2 m specific humidity, 10 m wind speed, and surface pressure were derived from the 37 year (1979–2015) 3-hourly 32 km output of the North American Regional Reanalysis (NARR) [*Mesinger et al*., 2006]. Additionally, NARR output fields of precipitation and shortwave radiation were utilized as "background" fields for their final NLDAS-2 counterparts derived as described later. All of these NARR outputs were interpolated to hourly temporal resolution and 0.125° spatial resolution (nominally 12 km). To account for elevation differences between the NLDAS grid and the NARR grid, a terrain height adjustment is applied to the air temperature and surface pressure using a uniform lapse rate of 6.5 k/km. The specific humidity and downward longwave radiation are then also adjusted based on the adjusted air temperature [see *Cosgrove et al*., 2003], with the original relative humidity being preserved.

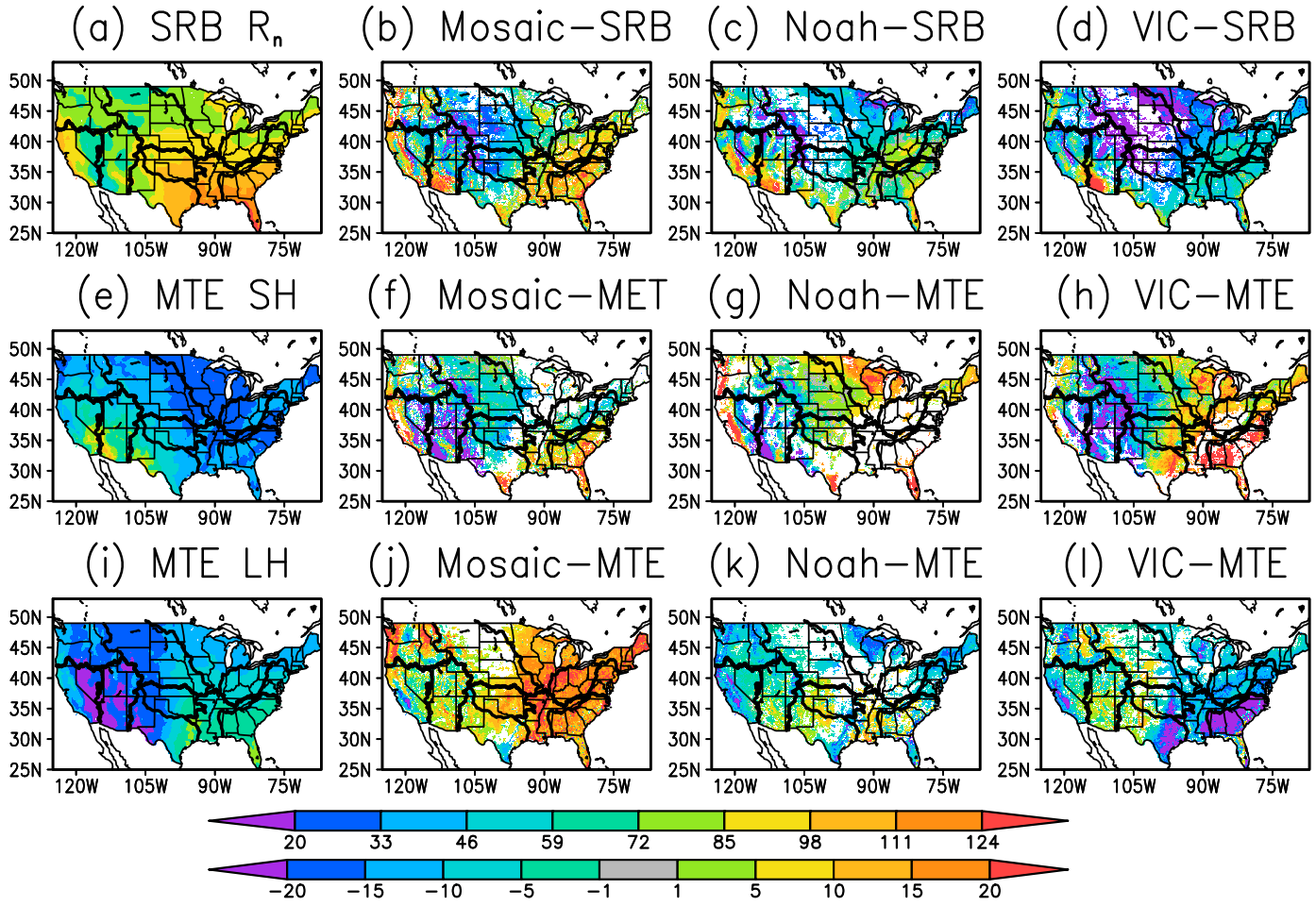

Figure 4. Spatial distribution of (a–d) mean annual net radiation (R_n), (e–h) sensible heat flux (SH), and (i–l) latent heat flux (LH) for the observed (Figures 4a, 4e, and 4i) and differences of simulations from for the operational NLDAS-2 simulated by Mosaic (Figures 4b, 4f, and 4j), Noah (Figures 4c, 4g, and 4k), and VIC (Figures 4d, 4h, and 4l) and the observations for the 24.5 year period of July 1983 to December 2007. Reference values are from MTE, except R_n from SRB (unit: W m^{-2}). The region that is not significant at 95% confidence level (Student's *t* test) is represented in white color. The first color bar is for mean annual value (Figures 4a, 4e, and 4i). The second color bar is for the differences (Figures 4b, 4f, and 4j; Figures 4c, 4g, and 4k; and Figures 4d, 4h, and 4l).

The NLDAS-2 precipitation forcing over the CONUS is anchored to the Climate Prediction Center (CPC) 0.125° gauge-only daily precipitation analysis with bias correction from monthly Precipitation-elevation Relationships on Independent Slope Model [*Daly et al.*, 1994] precipitation values. The NLDAS-2 daily precipitation is then temporally disaggregated to hourly values by applying hourly temporal weights derived from either EMC hourly stage II radar precipitation [*Lin and Mitchell*, 2005], CPC hourly CMORPH (CPC MORPHing technique) precipitation [*Joyce et al.*, 2004], the CPC 2 × 2.5° hourly precipitation analysis, or hourly NARR precipitation (in priority order, based on data availability). The NARR downward shortwave radiation is adjusted by the University of Maryland's Surface Radiation Budget (SRB) data set [*Pinker et al.*, 2003] through a ratio-based [*Berg et al.*, 2003] bias correction technique (see http://www.emc.ncep.noaa.gov/mmb/nldas/LDAS8th/forcing/forcing_narr.shtml). More details for NLDAS-2 forcing generation can be found in *Cosgrove et al.* [2003] and *Xia et al.* [2012].

2.3. NLDAS-2 Products and RFC Basin Mask Data

In this surface energy budget study we evaluate both of the NLDAS-2 primary radiative forcings (downward shortwave and longwave radiation) as well as the counterpart upward and net shortwave and longwave radiative fluxes simulated by the LSMs. Additionally, we evaluate the surface sensible and latent heat fluxes simulated by the LSMs. This evaluation is performed over the CONUS for the three NLDAS-2 operational LSMs that simulate the surface energy budget (Mosaic, Noah 2.8, and VIC4.0.3) and for the NLDAS-2 research versions of Noah and VIC (Noah-I and VIC4.0.5) for the 24.5 year period (1 July 1983 to 31 December 2007).

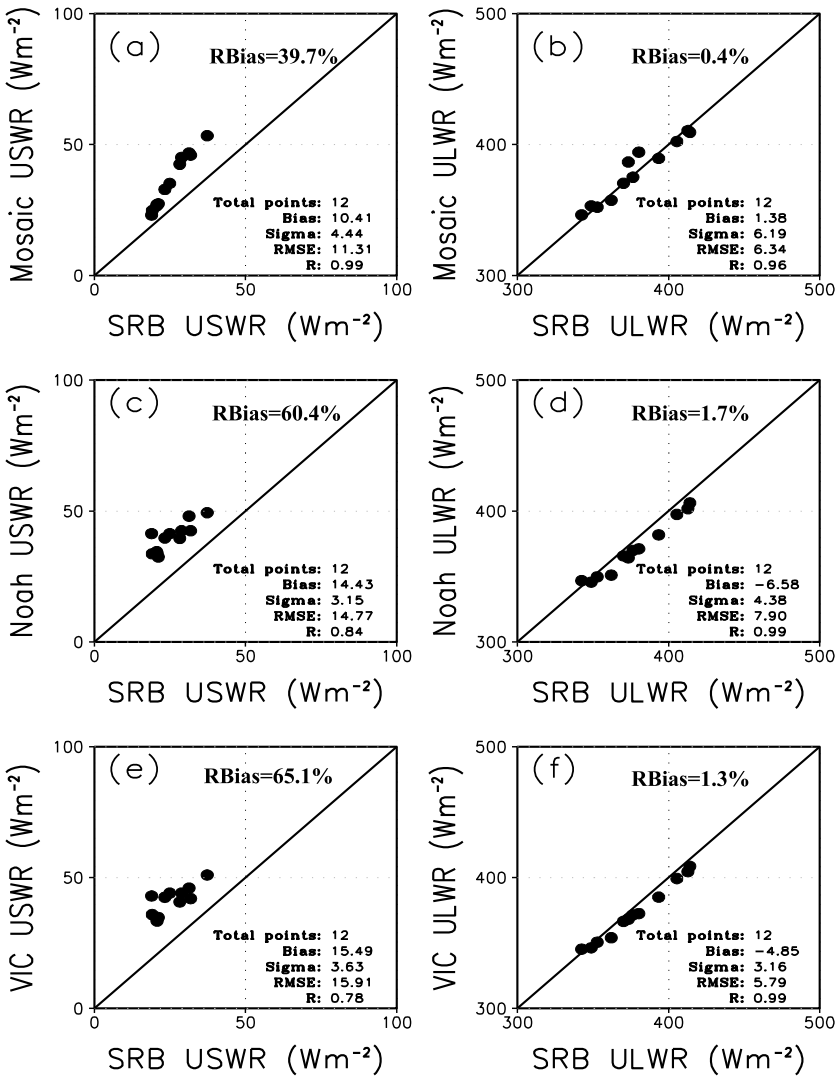

Figure 5. Scatterplots and statistical metrics for mean annual (a, c, and e) upward shortwave radiation (USWR) and (b, d, and f) upward longwave radiation (ULWR) of SRB derived versus operational NLDAS-2 simulated by Mosaic (Figures 5a and 5b), Noah (Figures 5c and 5d), and VIC (Figures 5e and 5f) for the 12 CONUS RFCs for the 24.5 year period of July 1983 to December 2007 (unit: W m^{-2}).

Fields of monthly mean energy fluxes are calculated from the hourly NLDAS-2 fields. The NLDAS-2 land-sea mask, inland lake mask, and RFC basin masks for each of the 12 CONUS RFCs (Figure 1) as used in NLDAS-1 [*Mitchell et al*., 2004] are also used in this study. The names and abbreviations of the 12 CONUS RFCs are listed in Table 2, which also shows the basin average 24.5 year mean annual Bowen ratio (ratio of sensible to latent heat flux) calculated from the observation-based sensible and latent heat flux fields of the MPI MTE FLUXNET database. For reference, in Table 2 we also provide the 30 year (1961–1990) mean annual basin average values of the so-called climate index, which is a ratio of precipitation (P) to potential evaporation (PE) (adapted from *Schaake et al*. [2004]). The results show that the driest RFCs are CBRFC and CNRFC, and the wettest RFCs are OHRFC and LMRFC. To present RFC-specific table entries and figure frames in a consistent order throughout this paper, all such tables and figures will follow the order of RFCs listed in Table 2 (from the driest CBRFC to the wettest OHRFC).

2.4. Reference Data Sets

The independent reference data sets used here for assessment include (1) monthly surface radiation data and (2) monthly sensible and latent heat flux data.

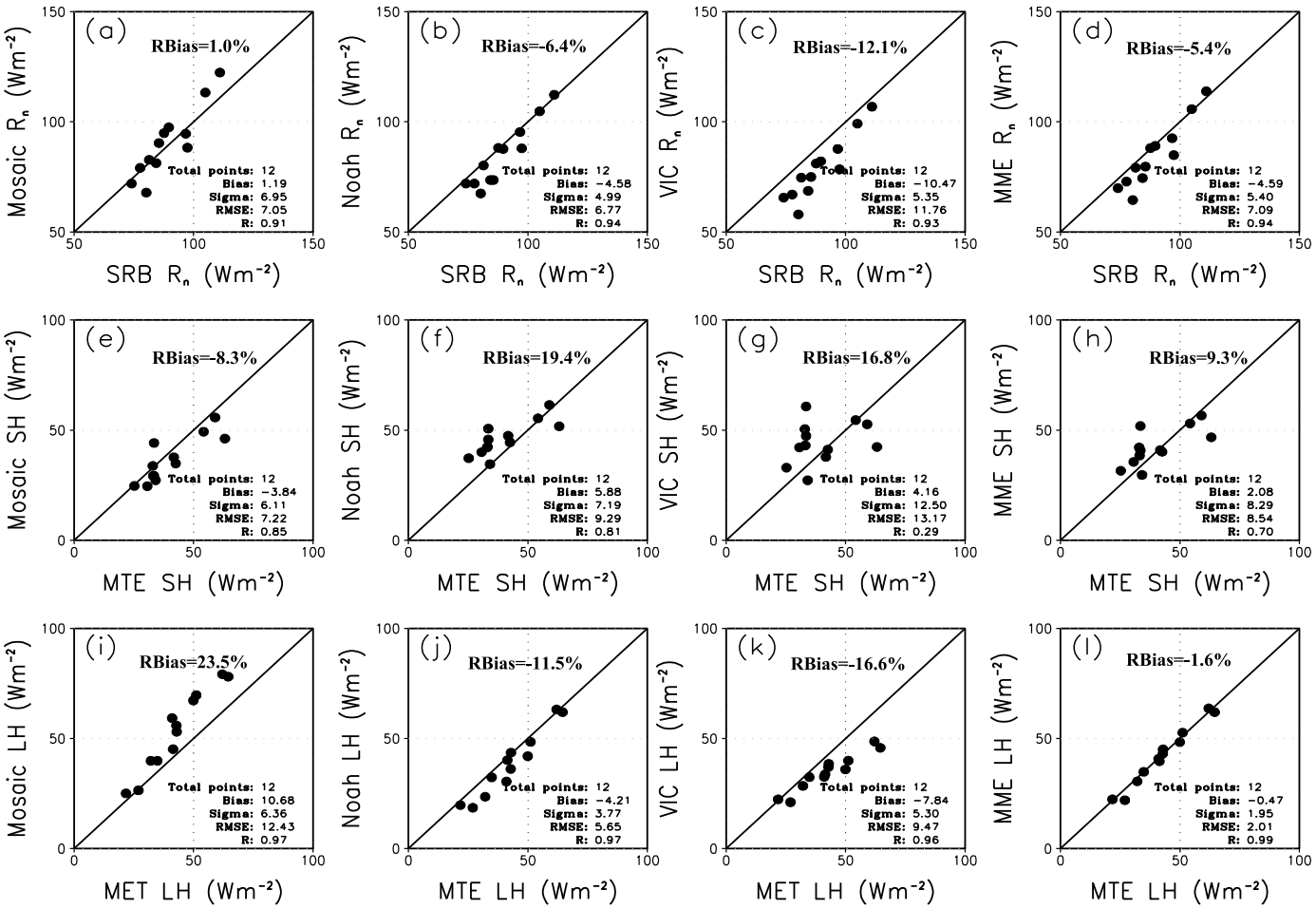

Figure 6. Scatterplots and their statistical metrics for (a–d) mean annual net radiation (R_n), (e–h) sensible heat flux (SH), and (i–l) latent heat flux (LH) of reference versus operational NLDAS-2 simulated by, from left to right, Mosaic, Noah, VIC, and their multimodel ensemble mean (MME) for the 12 CONUS RFCs for the 24.5 year period of July 1983 to December 2007. Reference SH and LH from MTE and reference R_n from SRB (unit: W m^{-2}).

The surface radiation data are from NASA GEWEX/SRB satellite observations (https://eosweb.larc.nasa.gov/project/srb/srb_table). The uncertainties of the GEWEX/SRB product with respect to the operational Baseline Surface Radiation Network (BSRN) [*Ohmura et al.*, 1998] measurements during the 1992–2007 period are believed to be about ±5.0 W m^{-2} for longwave radiation and ±5.0–20.0 W m^{-2} for shortwave radiation (also see http://gewex-srb.larc.nasa.gov/common/php/SRB_validation.php). The corresponding bias and root-mean-square error are 0.2 W m^{-2} and 11.1 W m^{-2} for longwave radiation and −4.3 W m^{-2} and 23.1 W m^{-2} for shortwave radiation, respectively [*Zhang et al.*, 2013, 2014]. Thus, mean bias results are within or nearly within the uncertainty for BSRN measurements [*Hinkelman et al.*, 2009].

By comparison, in situ flux measurement uncertainty is now estimated to be ±2.0 W m^{-2} for downward longwave radiation and ±4.0 W m^{-2} for downward shortwave radiation at monthly time scales [*Raschke et al.*, 2012]. However, in situ data were generally measured at individual stations and are only representative of small local areas around the points where the measurements are being made. Hence, they are difficult to compare against model simulations of a grid box (of nominal 144 km^2 area in this case) owing to local point versus grid box area scale differences [*Robock et al.*, 2003]. Furthermore, in situ measurements tend to be sparse in space and sometimes also in time. Although the SRB satellite-based data sets have slightly larger uncertainty than the in situ measurements, the good spatial coverage and long temporal duration of the SRB data sets make them a favorable alternative to in situ measurements. As a result, we believe the SRB data set is appropriate to evaluate the reanalysis products generated from land surface models and climate models [*Liang et al.*, 2007; *Yang et al.*, 2015].

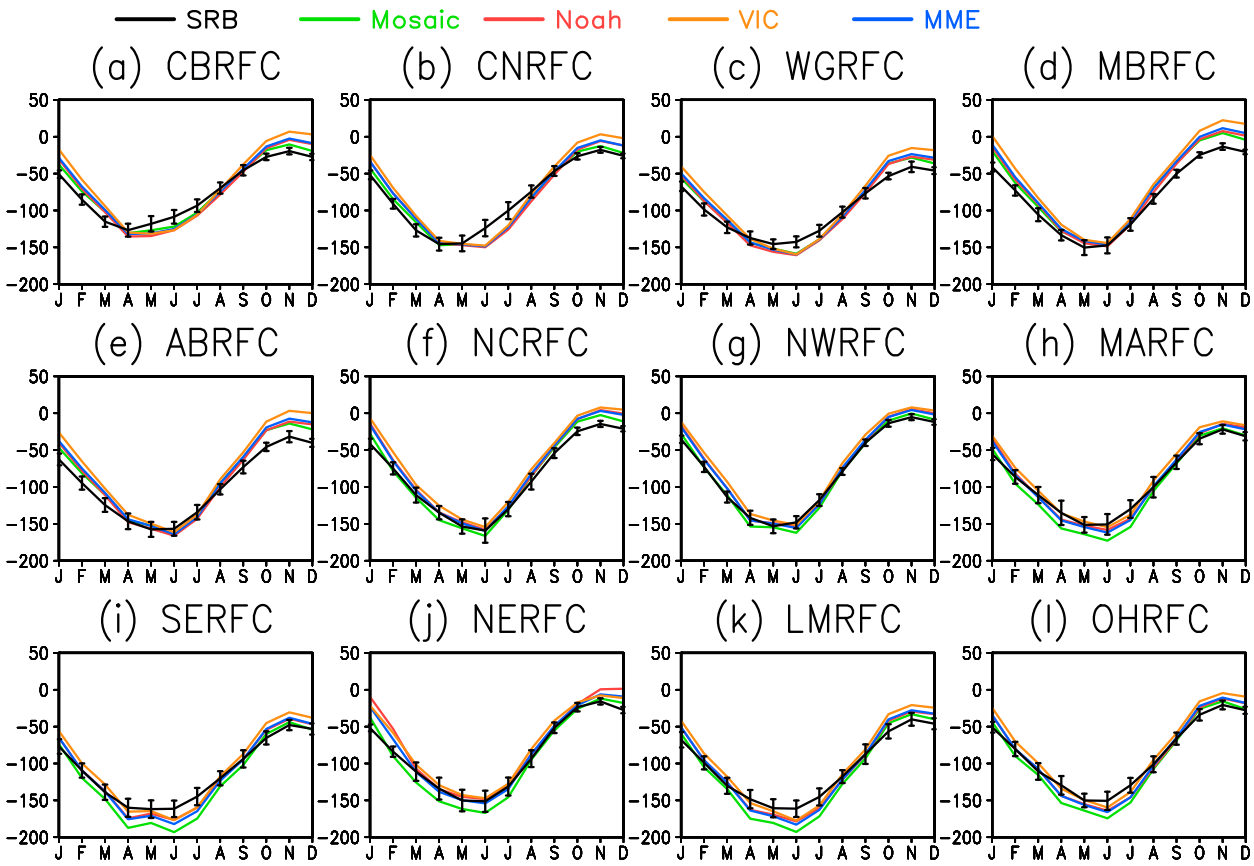

Figure 7. Mean annual cycle of monthly mean net radiation (R_n) and its error bar (unit: W m^{-2}) from SRB derived (black) and operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007. The error bars can represent interannual variability.

Globally gridded estimates of land surface sensible and latent heat flux data at 0.5° spatial resolution on a monthly time scale are derived and available from the Max Planck Institute (MPI), Germany [*Jung et al.*, 2009, 2010]. The latter are referred to as a multitree ensemble (MTE) product because they are calculated by an MTE method that applies (1) a "machine learning" approach for the upscaling of eddy covariance measurements at FLUXNET sites [*Baldocchi et al.*, 2001], (2) meteorological data, (3) climate data, and (4) the fraction of absorbed photosynthetically active radiation data. For consistency with the period of the NASA GEWEX/SRB data set described earlier, we also use the period from July 1983 to December 2007 for our application of the MPI MTE-based sensible and latent heat flux fields. The MPI sensible and latent heat flux data sets have been used as key benchmarking data sets by many studies [*Jung et al.*, 2010; *Peters-Lidard et al.*, 2011; *Bentsen et al.*, 2013; *Stegehuis et al.*, 2013; *Velpuri et al.*, 2013; *Cai et al.*, 2014].

In order to use the NLDAS-2 0.125° spatial mask data set to calculate basin average values for the 12 RFCs, the monthly GEWEX/SRB radiation data and MTE sensible/latent heat flux data are regridded to the 0.125° NLDAS grid over CONUS using the water budget conservation method described in *Xia et al.* [2015a]. This approach will conserve the interpolated values when they are aggregated to the original grid so that the effect of interpolation errors is minimal for RFC basin averages.

3. Evaluation of Surface Energy Budget Components

The monthly surface energy budget can be written as

$$R_n - \mathrm{SH} - \mathrm{LH} - G = 0 \tag{1}$$

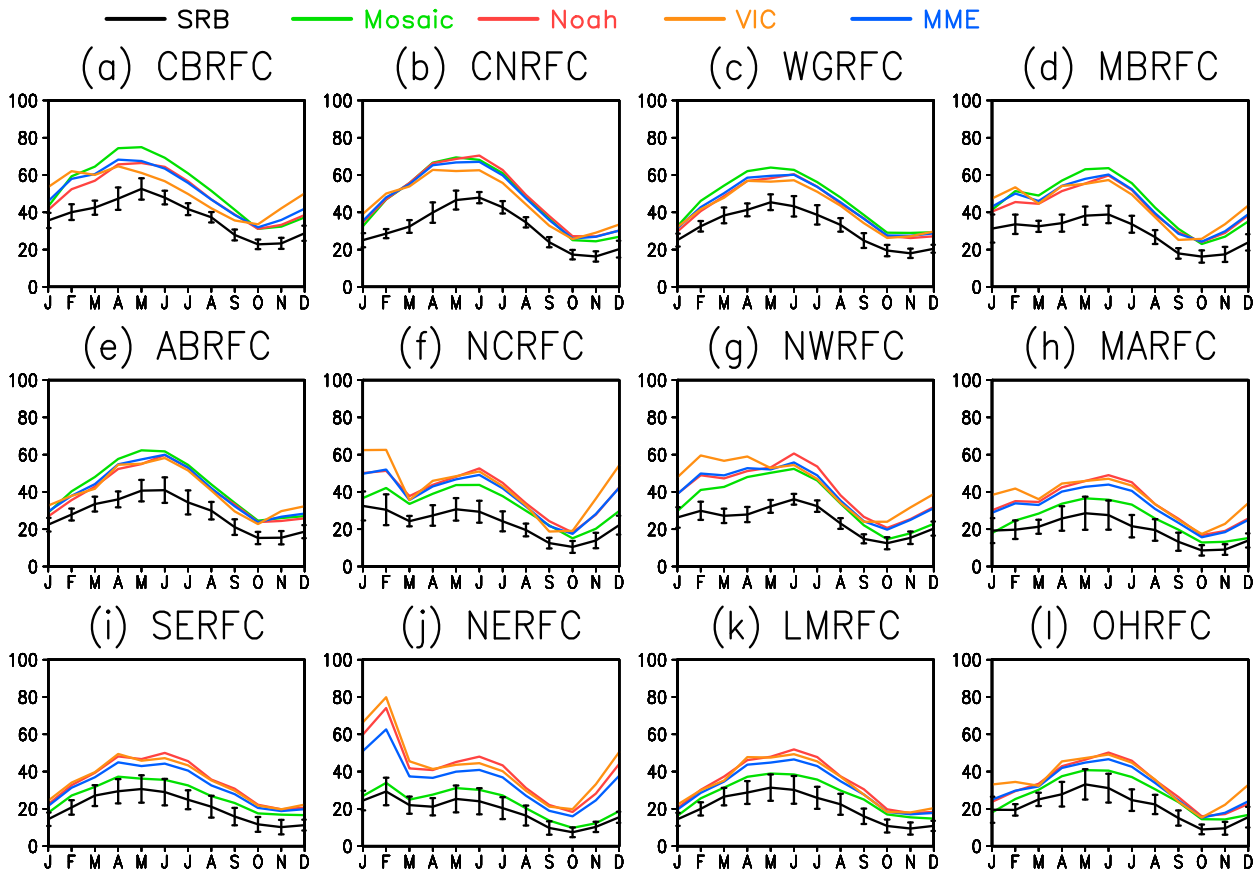

Figure 8. Mean annual cycle of monthly mean upward shortwave radiation (USWR) and its error bar (unit: W m^{-2}) from the SRB derived (black) and operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007.

in which

$$R_n = \mathrm{DSWR} - \mathrm{USWR} + \mathrm{DLWR} - \mathrm{ULWR} \tag{2}$$

where R_n is the surface net radiation and DSWR, USWR, DLWR, and ULWR are the surface downward shortwave radiation, surface upward shortwave radiation, surface downward longwave radiation, and surface upward longwave radiation, respectively. Additionally, SH is the sensible heat flux, LH is the latent heat flux, and G is the ground heat flux.

In these models (i.e., Mosaic, Noah, and VIC), the sensible heat flux SH is calculated from a typical bulk transfer formulation given by

$$\mathrm{SH} = -\rho C_p \mathrm{CH}|V|(T_a - T_{\mathrm{skin}}) \tag{3}$$

where ρ is the air density, C_p is the heat capacity for air, T_a is the air temperature, T_{skin} is the land surface skin temperature, and CH is the surface turbulent exchange coefficient for heat. The product CH $|V|$ is the aerodynamic conductance and its reciprocal is the aerodynamic resistance. In equation (3), a positive SH indicates a heat source to the atmosphere and heat sink to the land surface when T_{skin} exceeds T_a. A negative SH value indicates a heat sink to the atmosphere and heat source to the land surface when T_{skin} is below T_a. The different land surface models have different CH formulations and T_{skin} values.

The latent heat flux LH is calculated as

$$\mathrm{LH} = L_v E_d + L_v E_c + L_v E_t + L_s E_s \tag{4}$$

where E_d is the direct evaporation from bare soil, E_c is the evaporation of precipitation intercepted by the vegetation canopy, E_t is the transpiration from the vegetation canopy and roots, and E_s is sublimation from the snowpack. L_v is the specific latent heat for vaporization, and L_s is specific latent heat for sublimation. In the Noah and

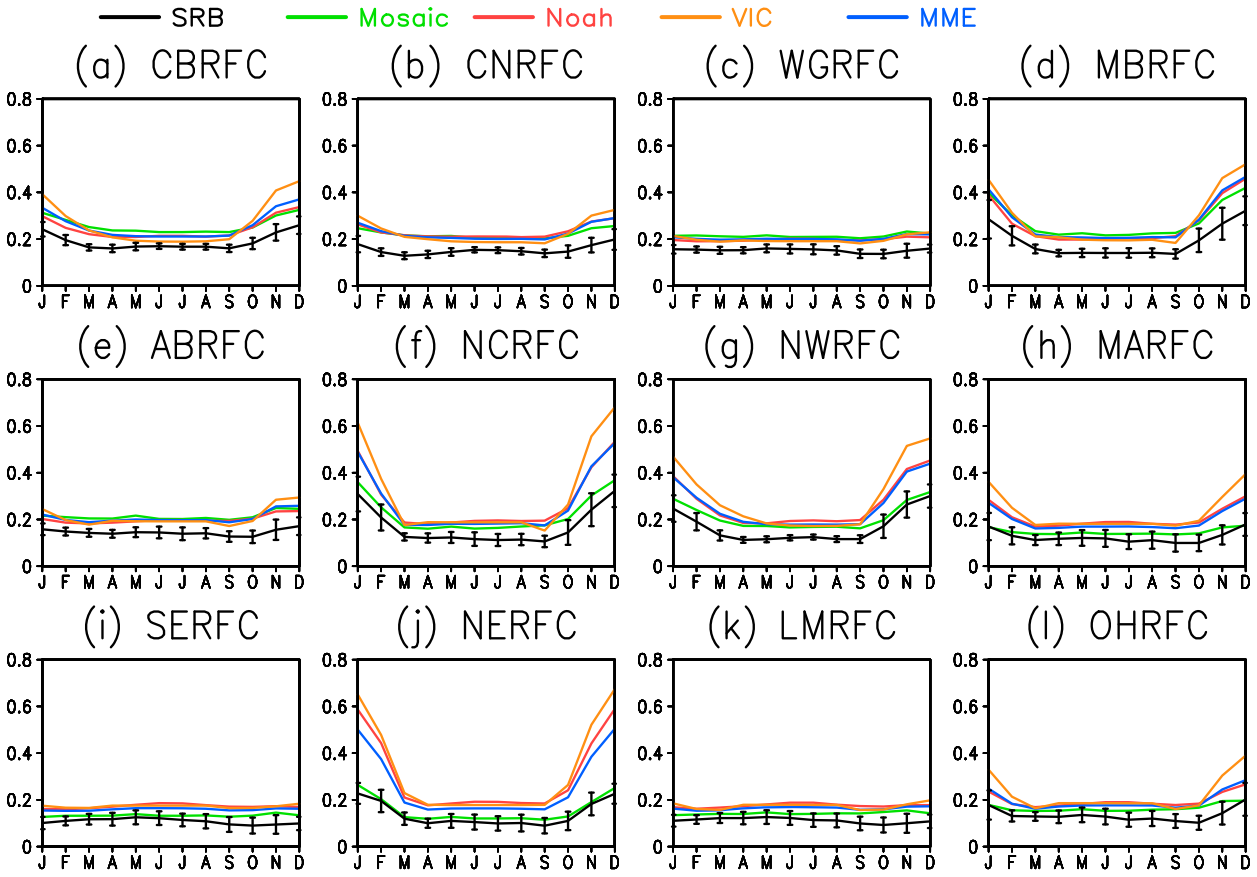

Figure 9. Mean annual cycle of monthly mean derived albedo (USWR/DSWR) and its error bar from the SRB derived (black) and operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007.

VIC models these terms depend directly on the calculation of potential evaporation (PE), which in the VIC model is obtained from the Penman-Monteith equation [*Monteith*, 1965] and in the Noah model is obtained from the stability-dependent revised form of the Penman equation presented and evaluated by *Mahrt and Ek* [1984]. The Mosaic model follows the simple biosphere model of *Sellers et al.* [1986] but is simplified enough to be written in Penman-Monteith form [*Koster and Suarez*, 1992]. In all three LSMs, E_d is a function of the soil moisture content [*Xia et al.*, 2015b] of the topmost soil layer and PE, E_c is a function of the canopy water content and PE, E_t is a function of canopy/surface conductance and PE, and E_s is a function of the snow water equivalent content of the snowpack and PE. It should be noted that the specific functions for the four evaporation components in equation (4) vary from model to model. More details about these functions can be found in *Koster and Suarez* [1996] for the Mosaic model, *Ek et al.* [2003] for the Noah model, and *Liang et al.* [1994] for the VIC model.

The ground heat flux G is calculated as

$$G = \left(\frac{K_T}{\Delta Z}\right)\left(T_{\text{skin}} - T_{\text{soil}}\right) \tag{5}$$

where K_T is the soil thermal conduction varying with soil type and soil moisture, ΔZ is the topsoil layer thickness, T_{skin} is the surface skin temperature, and T_{soil} is the topsoil layer temperature. This heat conduction equation is used in the Noah and VIC models. In the Mosaic model, the ground heat flux G is calculated as

$$G = \frac{DC}{\sqrt{2}}\frac{\mathrm{d}T_{\text{sfc}}}{\mathrm{d}t} - \frac{\omega DC}{\sqrt{2}}\left(\overline{T} - T_{\text{skin}}\right) \tag{6}$$

where D is the depth over which a diurnal temperature wave is felt, C is the volumetric heat capacity, ω is the frequency of the diurnal temperature cycle, and $\overline{T}$ is the temperature below the surface that is unaffected by the diurnal temperature cycle. The force restore equation (6) is consistent with the standard force restore formulation of *Deardorff* [1978].

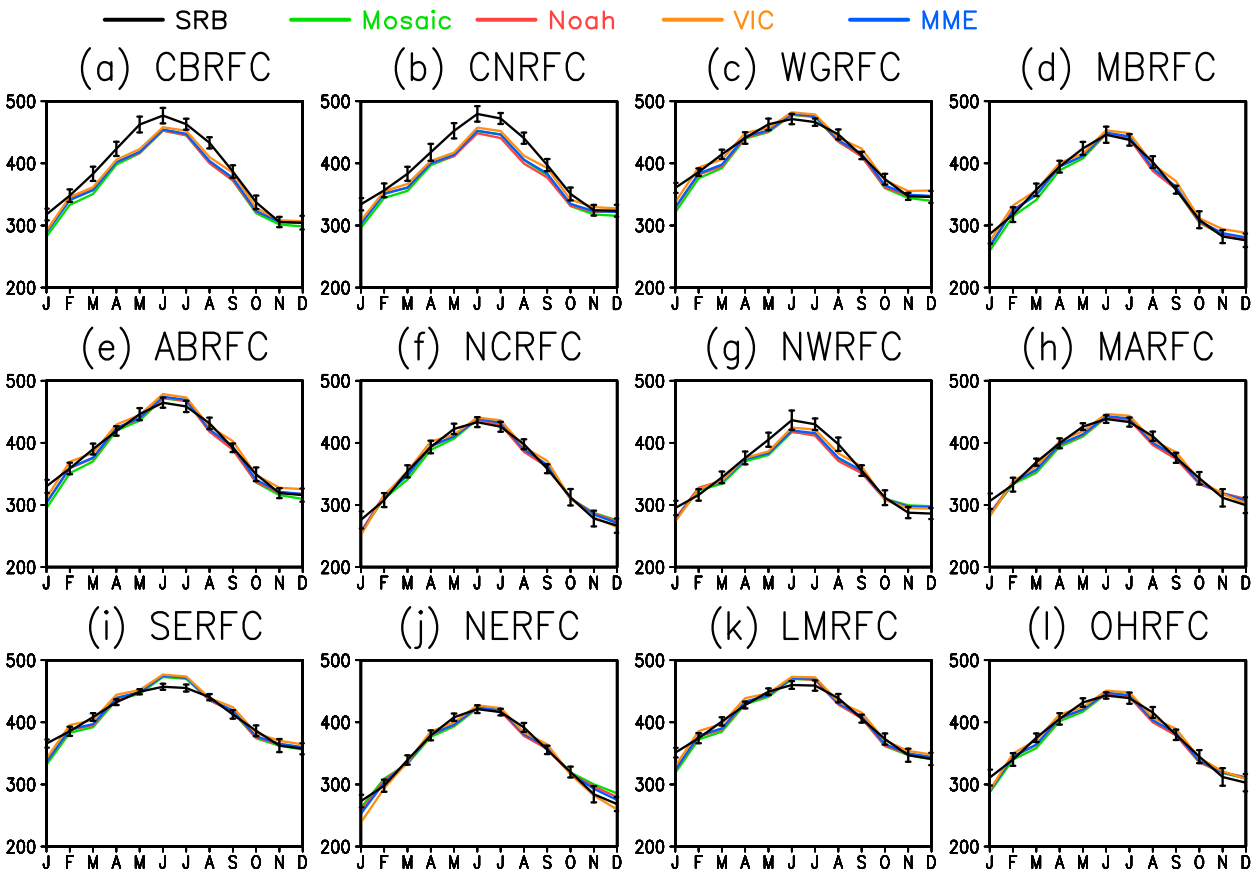

Figure 10. Mean annual cycle of monthly mean upward longwave radiation (ULWR) and its error bar (unit: W m^{-2}) from the SRB derived (black) and operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007.

Upward shortwave (USWR) and longwave (ULWR) radiation are expressed as

$$\mathrm{USWR} = \alpha_{\mathrm{ave}}\mathrm{DSWR} \tag{7}$$

$$\mathrm{ULWR} = \varepsilon\sigma T_{\mathrm{skin}}^4 + (1-\varepsilon)\mathrm{DLWR} \tag{8}$$

where α_{ave} is the average surface albedo for a given grid box, ε is the land surface emissivity, T_{skin} is land surface skin temperature, and σ is the Stefan–Boltzmann constant. It should be noted that the representation of upward shortwave radiation given by equation (7) is applied in the Noah and Mosaic models. The VIC model includes formulations for the absorption of shortwave radiation by the vegetation canopy. More specifically in VIC, the average albedo and upward shortwave radiation depend on multiple reflection and absorption processes between the canopy and the ground. In all three LSMs (Noah, Mosaic, and VIC), the infrared reflectance term (second term in equation (8)) is set to zero. In the absence of snow, all three models use an emissivity of 1 and the infrared reflectance term is indeed zero. With snow cover present, the effect of snow on the emissivity is ignored in all three LSMs, as the emissivity value remains 1.

Table 3 lists the 24.5 year mean annual energy budget components calculated from the SRB and MTE data sets. The energy balance closure across these terms is reasonable. Six RFCs have closure errors within 10%, and 11 of the 12 RFCs are within a 14% imbalance. The NCRFC shows the largest energy closure imbalance of 19%. Given the uncertainties of MTE sensible and latent heat flux and the lack of a ground heat flux term ($<$3.0 W m^{-2}) in the closure calculations, the imbalance errors are reasonable, providing confidence in the use of these reference data sets for the evaluation of both the NCEP operational and research NLDAS-2 systems.

Throughout our evaluations of the surface energy fluxes, we apply the metrics of root-mean-square error, bias, relative bias (RBias), Sigma, correlation (*R*), Taylor skill score *S* [*Taylor*, 2001], and monthly anomaly correlation (AC). For detailed definitions of these metrics, see our supporting information (Text S1), which also defines the significance test we apply to the difference between the model simulated and SRB or MTE reference values.

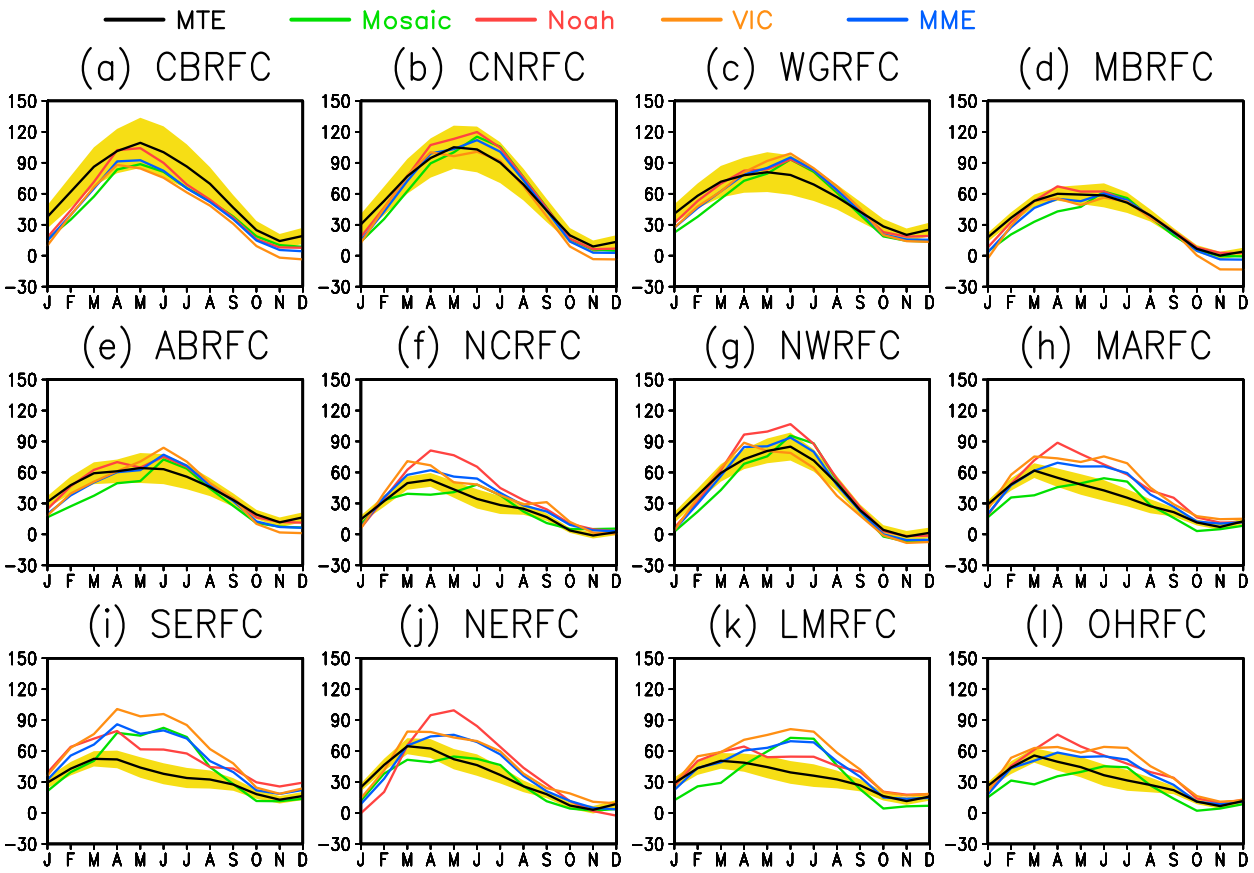

Figure 11. Mean annual cycle of monthly mean sensible heat flux (SH) (unit: W m^{-2}) from the MTE reference (black) and from the operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007. Yellow shading represents the spread of the FLUXNET-based multipletree ensemble (MTE).

3.1. Evaluation of NLDAS-2 Downward Shortwave and Longwave Radiation

The first step in our study is to use the GEWEX/SRB satellite-based reference values to evaluate the primary NLDAS-2 radiative forcing fields, which are the downward shortwave and longwave radiation. In Figure 2, the spatial distribution of mean annual values of DSWR (a and b) and DLWR (c and d) for both SRB (Figures 2a, 2c, and 2e) and NLDAS-2 (Figures 2b, 2d, and 2f) shows that SRB and NLDAS-2 DSWR are quite similar, except that the SRB has smaller DSWR values in higher latitudes (Figures 2a and 2b). The NLDAS-2 field of DLWR manifests the obvious elevation-adjustment characteristics in western mountainous regions (Figure 2d), and SRB does not (Figure 2c). The relative bias is generally in range from −1.0% to 10.0% for DSWR (Figure 2e) and in range from −5.0% to 1.0% for DLWR over most regions of CONUS except for mountainous regions (Figure 2f). The scatterplots in Figure 3 for the 12 RFCs show that although NLDAS-2 slightly overestimates (underestimates) DSWR (DLWR) compared to SRB, the bias is very small (9.5 W m^{-2} and −5.7 W m^{-2}, respectively). This suggests that the quality of NLDAS-2 radiation forcing is quite good in this mean annual aggregate comparison. The overestimation of DSWR in NLDAS-2 mainly comes from the warm season at 12 RFCs, in particular at OHRFC, MARFC, LMRFC, and SERFC. In contrast to DSWR, the underestimation of DLWR in NLDAS-2 mainly appears in the cool season at 12 RFCs. In these aggregate comparisons, NLDAS-2-simulated radiation forcing and SRB reference data set show fairly good agreement as most simulation values are within the SRB uncertainty range (not shown).

3.2. Evaluation of Operational NLDAS-2 Energy Budget Components

The annual means and climatologically averaged seasonal cycles simulated by Mosaic, Noah, VIC, and their ensemble mean MME are compared against SRB and MTE fluxes in the following section. For MME, it should

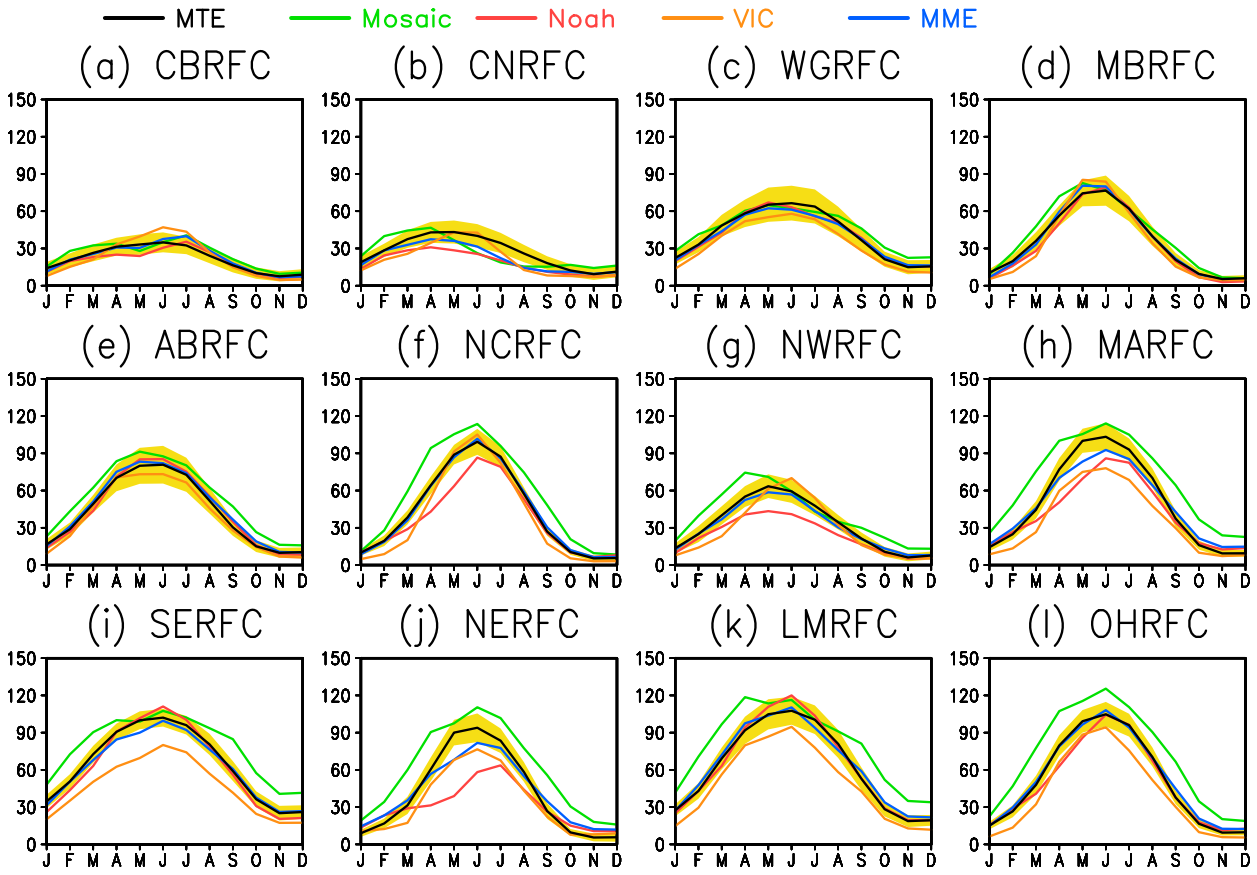

Figure 12. Mean annual cycle of monthly mean latent heat flux (LH) (unit: W m^{-2}) from the MTE reference (black) and from the operational NLDAS-2 simulated by Mosaic (green), Noah (red), VIC (orange), and MME (blue) for each CONUS RFC for the 24.5 year period of July 1983 to December 2007. Yellow shading represents the spread of the multiple tree ensemble (MTE).

be interpreted with caution because this analysis is based on a very small sample of land surface models (only three models), so the sample size is definitely not enough to demonstrate the merit of central limit theorem, although MME does improve both the correlation and relative bias for the latent heat flux simulation. As more land surface models are incorporated into the current NLDAS-2 and future NLDAS-3 projects, we will report that more robust MME results with larger sample size (e.g., of order 10 models).

3.2.1. Comparison of Annual Means

For the surface energy budget, the LSMs partition the net incoming radiation into sensible heat, latent heat, and ground heat flux terms. In general, the annual mean ground heat flux is very small ($<$3.0 W m^{-2}) and therefore can be considered negligible at the annual time scale. Figure 4 shows a comparison of annual mean net radiation, sensible heat flux, and latent heat flux from SRB, MTE, and the three NCEP operational NLDAS-2 LSMs with energy cycles (Mosaic, Noah, and VIC). The three LSMs basically capture the spatial distribution of SRB net radiation, except that the LSMs have smaller net radiation than SRB in the interior CONUS. The VIC model has the smallest net radiation when compared with the Mosaic and Noah models (Figures 4a–4d). The Mosaic model generally underestimates (overestimates) sensible (latent) heat flux estimates relative to the MTE data set, in particular over the eastern U.S. (Figures 4e, 4f, 4i, and 4j). In contrast, the Noah and VIC models generally overestimate (underestimate) sensible (latent) heat flux estimates compared to MTE (Figures 4g, 4h, 4k, and 4l). Over the Northern Great Plains, the difference between the simulations and MTE reference values is not significant for net radiation and latent heat flux at the 95% confidence level (Student's *t* test). This is also evident over portions of Southern Great Plains for both sensible and latent heat fluxes. For Noah, there is no significant bias in the sensible heat flux simulation over the south and southeast, suggesting a good performance. Since all three LSMs use the same incoming radiation forcing, which is close to the SRB reference values (Figures 2 and 3), the main differences between SRB and NLDAS-2 net radiation

Table 4. Anomaly Correlation (AC) Between SRB or MTE Reference Values and Model-Simulated Monthly Net Radiation (R_n), Sensible Heat Flux (SH), and Latent Heat Flux (LH) From the NCEP Operational NLDAS-2 as Calculated for the 12 CONUS RFCs Over the 24.5 year Period (July 1983 to December 2007)[a]

	R_n				SH				LH			
RFC Label	Mosaic	Noah	VIC	MME	Mosaic	Noah	VIC	MME	Mosaic	Noah	VIC	MME
CBRFC	**0.29**	0.24	0.28	**0.29**	**0.66**	0.52	0.21	0.56	0.80	0.76	0.78	**0.83**
CNRFC	**0.21**	*0.10*	0.18	0.18	**0.73**	0.67	0.21	0.63	0.81	0.80	0.65	**0.82**
WGRFC	0.04	**0.11**	*−0.06*	*0.03*	0.81	**0.84**	0.56	0.80	0.80	0.82	0.80	**0.83**
MBRFC	0.22	**0.35**	0.32	0.33	**0.62**	0.59	0.41	0.61	0.74	0.77	0.80	**0.83**
ABRFC	0.22	**0.28**	0.13	0.24	0.69	**0.73**	0.35	0.66	0.72	0.77	0.71	**0.78**
NCRFC	0.40	**0.52**	0.42	0.47	**0.49**	0.33	0.21	0.42	0.40	0.48	**0.61**	0.58
NWRFC	0.27	0.28	0.25	**0.29**	**0.69**	0.57	*0.03*	0.56	0.59	0.58	0.63	**0.69**
MARFC	0.51	**0.54**	0.49	**0.54**	0.35	0.34	0.27	**0.36**	0.22	**0.37**	0.12	0.29
SERFC	0.54	**0.59**	0.51	0.56	0.65	**0.73**	0.62	0.68	*−0.07*	**0.28**	*0.08*	*0.07*
NERFC	0.55	0.52	0.54	**0.56**	**0.29**	*0.10*	*0.09*	0.18	0.36	0.21	**0.50**	0.49
LMRFC	0.48	**0.54**	0.44	0.50	0.61	**0.70**	0.54	0.64	*0.05*	**0.30**	0.13	0.15
OHRFC	0.46	**0.54**	0.44	0.50	**0.48**	0.46	0.33	**0.48**	0.39	**0.48**	0.36	**0.48**

[a]Bold font denotes the maximum value for each variable and basin. Italic font denotes an insignificant anomaly correlation (<0.12) at the 5% significance level (Student's *t* test). Reference R_n is from the SRB/GEWEX satellite estimates. References SH and LH are from the FLUXNET-based MTE method. The MME is the three-model ensemble mean.

are from the LSM-simulated upward shortwave and upward longwave radiation fields. An analysis at each of the 12 RFCs (Figure 5) shows that all three LSMs overestimate SRB upward shortwave radiation (relative biases are ~40%–65%), but closely capture upward longwave radiation (relative bias is in range from 0.4% to 1.7%), when compared to SRB reference values. Overall, two of the three LSMs plus the three-model ensemble mean (MME) underestimate SRB net radiation (Figure 6), mostly due to the overly large upward shortwave radiation and the somewhat too small upward longwave radiation (Figure 5). The three LSMs either underestimate or overestimate sensible and latent heat flux compared to the MTE FLUXNET-based reference data set, and notably, their ensemble mean is closest to the MTE results (Figure 6). The relative biases vary from −12.1% for VIC and −6.4% for Noah to 1.0% for Mosaic when R_n is compared, from −8.3% for Mosaic to 16.8% for VIC and 19.4 for Noah when SH is compared, and from −16.6% for VIC and −11.5% for Noah to 23.5% for Mosaic when LH is compared. MME has the smallest relative bias for LH (−1.6%). VIC has a very low correlation for its sensible heat flux simulation with respect to the MTE reference product, due to its overly large sensible flux values in the eastern CONUS (Figure 4).

Table 5. For the 12 CONUS RFCs Over the 24.5 year Period (July 1983 to December 2007), Anomaly Correlation (AC) Between SRB Reference and NLDAS-2 Monthly Downward Shortwave (DSWR) and Longwave (DLWR) Radiation and Between SRB Reference and Operational NLDAS-2 Model-Simulated Monthly Upward Shortwave (USWR) and Longwave (ULWR) Radiation[a]

	Forcing		USWR				ULWR			
RFC Label	DSWR	DLWR	Mosaic	Noah	VIC	MME	Mosaic	Noah	VIC	MME
CBRFC	0.78	0.86	0.38	**0.50**	0.42	0.47	0.67	0.66	**0.77**	0.72
CNRFC	0.82	0.88	0.35	**0.52**	0.41	0.47	0.76	0.77	**0.89**	0.81
WGRFC	0.87	0.85	0.20	**0.37**	0.28	0.30	0.79	0.74	**0.85**	0.80
MBRFC	0.81	0.79	0.47	**0.66**	0.63	0.63	0.89	0.91	**0.93**	0.92
ABRFC	0.85	0.78	0.16	**0.28**	0.21	0.24	0.86	0.86	**0.91**	0.89
NCRFC	0.82	0.80	0.44	0.52	**0.53**	0.52	0.89	0.90	**0.92**	0.91
NWRFC	0.86	0.83	0.48	**0.59**	0.55	0.58	0.79	0.79	**0.82**	0.81
MARFC	0.84	0.83	0.24	0.32	**0.34**	0.33	0.88	0.89	**0.92**	0.91
SERFC	0.87	0.93	*0.00*	*0.00*	*−0.01*	*0.00*	0.91	0.90	**0.93**	0.92
NERFC	0.83	0.83	0.23	0.32	**0.37**	0.34	0.88	0.88	0.88	**0.89**
LMRFC	0.83	0.89	*0.03*	*0.01*	*0.04*	*0.03*	0.89	0.89	**0.92**	0.91
OHRFC	0.83	0.85	0.24	0.34	**0.39**	0.37	0.88	0.89	**0.92**	0.90

[a]Bold font denotes the maximum value for each variable and basin. Italic font denotes an insignificant anomaly correlation (<0.12) at 5% significance level (Student's *t* test).

Table 6. Correlation Between SRB Reference and Operational NLDAS-2 Model-Simulated Monthly USWR for the 12 CONUS RFCs

RFC Label	Mosaic	Noah	VIC	MME
CBRFC	0.91	0.92	0.77	0.92
CNRFC	0.93	0.94	0.90	0.93
WGRFC	0.92	0.92	0.92	0.92
MBRFC	0.87	0.90	0.86	0.90
ABRFC	0.87	0.87	0.83	0.87
NCRFC	0.80	0.83	0.77	0.83
NWRFC	0.87	0.92	0.81	0.90
MARFC	0.72	0.75	0.71	0.75
SERFC	0.78	0.78	0.78	0.78
NERFC	0.75	0.73	0.68	0.74
LMRFC	0.80	0.80	0.79	0.80
OHRFC	0.77	0.79	0.77	0.80

3.2.2. Mean Annual Cycle Comparison

In general, all the NLDAS-2-simulated radiation components show a clear seasonality at each of the 12 RFCs, as expected (Figures 7–10). The error bars for the observations are calculated as a standard deviation for a given month when the interannual variability is considered. The surface net radiation (Figure 7) follows its seasonal course stemming from the changing Sun-Earth geometry. Comparison of upward shortwave radiation (USWR) shows that all three LSMs overestimate USWR as compared to the SRB at all 12 RFCs, in particular for these cases outside of error bars (Figure 8). The reason is that all three have a positive bias in their integrated surface albedo (Figure 9), calculated from the ratio of upward shortwave radiation to downward incoming shortwave radiation (USWR/DSWR).

As discussed in *Xia et al*. [2012], Mosaic has a smaller positive bias in surface albedo values than either the Noah or VIC models. The simulation by Mosaic is closer to the SRB values at OHRFC, NERFC, MARFC, LMRFC, and SERFC when compared to the other two LSMs. The reason is that the albedo values assigned in Mosaic are smaller than the values assigned in Noah and VIC for both bare soil and with vegetation cover, even when the same soil texture, vegetation type, and leaf area index (LAI) data are used [*Mitchell et al*., 2004; *Xia et al*., 2012].

During the snow season the notably different snow cover fraction that is simulated by the three NLDAS LSMs (see more below) despite their common precipitation and air temperature forcing is most likely the primary cause of the differences in the integrated surface albedo [*Sheffield et al*., 2003; *Mitchell et al*., 2004]. In the LSMs over snow cover, the integrated albedo is calculated as the averaged albedo from the vegetation fraction (tiling for Mosaic and VIC, predominant for Noah) and the snow cover fraction. Over the four most northern RFCs (Figure 1), which have a longer season of generally persistent snow cover, the three LSMs and their ensemble mean (MME) have a larger-amplitude seasonal cycle of surface albedo, with distinctly higher simulated albedo as compared to the SRB-derived albedo during the winter months. The magnitudes of the integrated albedo are large and small in winter and summer, respectively, due to large snow albedo effects. In the NCEP operational NLDAS-2 system, the vegetation albedo, bare soil albedo, and maximum snow albedo are assigned parameters. In the three LSMs, upward shortwave radiation is calculated using different methods, such as using dominant vegetation type or tiling methods (see Table 1). Also, the three LSMs apply different algorithms for calculating the snow cover fraction as a function of snow depth and vegetation cover [*Sheffield et al*., 2003]. Altogether, these factors lead to different simulated albedo values across the LSMs. Since the simulated winter albedo in the northern RFCs is too large in all three LSMs when compared to SRB, we conclude that all three LSMs overamplify the effects of snow cover on the surface albedo. This is due to both the assumed maximum value of snow albedo for deep snow and the assignment of snow cover fraction as a function of snow depth and needs further investigation.

For the nonwinter seasons, the simulated integrated albedo is also larger than the SRB-derived values, also requiring a reassessment of the prescribed albedo parameters. Note that the integrated albedo from Mosaic is very close to the derived values from SRB at NERFC, MARFC, LMRFC, and SERFC. These RFCs will allow us to further investigate the albedo calculation in Mosaic, versus that in the other two LSMs. A key point, then, is that the albedo values assigned in the NLDAS-2 LSMs, both for snow cover and nonsnow covered but vegetated regions, need to be revisited and further evaluated against in situ observations and satellite-based estimates in order to attain more realistic and common albedo values among the LSMs employed in the operational and research NLDAS-2 versions.

The comparison of the mean annual cycle of model-simulated versus SRB reference upward longwave radiation (ULWR) for the 12 RFCs shows that all three LSMs and their ensemble mean in particular underestimate

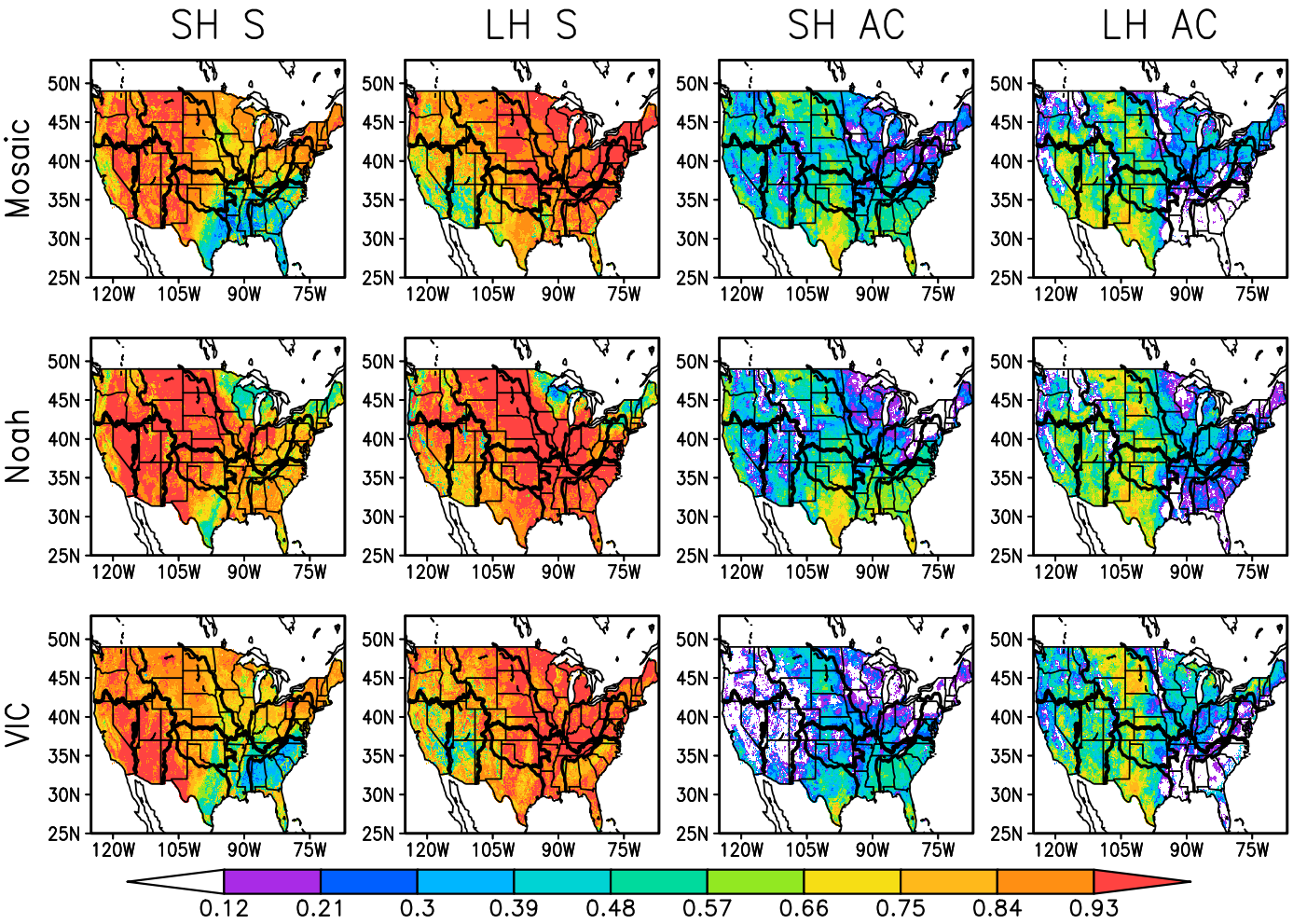

Figure 13. Spatial distribution of Taylor skill score (*S*) of (column 1) sensible heat flux (SH) and (column 2) latent heat flux (LH) and anomaly correlation score (AC) of (column 3) SH and (column 4) LH of operational NLDAS-2 simulations by (row 1) Mosaic, (row 2) Noah, and (row 3) VIC for the 24.5 year period of July 1983 to December 2007 calculated with respect to the monthly mean MTE observed values.

the SRB-derived ULWR values at CNRFC and CBRFC (Figure 10), suggesting either a low bias in the LSM skin temperature or surface emissivity or both (see equation (8)). Notably, these two RFCs are the most arid of the RFCs with the least amount of vegetation cover. It is more difficult to assign surface emissivity over sparsely vegetated land cover. For the 10 other RFCs, all three LSMs broadly capture the SRB seasonal cycle of ULWR, though the simulated monthly variability is larger than that of the SRB (Figure 10).

The seasonal cycle of sensible and latent heat flux (Figures 11 and 12) closely follows the seasonal course of the surface net radiation. The yellow shading area is the spread of MTE flux values provided in the original MTE products. If the simulation is within the yellow area, it is considered a reasonable simulation. Otherwise, it is considered as an overestimation or underestimation. In general, the Mosaic model underestimates (overestimates) MTE sensible (latent) heat flux for most of the 12 RFCs. In contrast, the Noah and VIC models overestimate (underestimate) MTE sensible (latent) heat flux for many RFCs, such as NWRFC, NCRFC, NERFC, and MARFC. All three models overestimate (underestimate) MTE sensible (latent) heat flux at LMRFC and SERFC, especially for the VIC model.

Besides the comparison of annual means and mean annual seasonal (monthly) cycles of energy fluxes given above, we also calculate monthly anomaly correlation (AC) to evaluate model simulation skill (Tables 4 and 5). For the R_n simulation results in Table 4, the Noah LSM is the most skillful, followed by MME, Mosaic, and VIC, respectively. For the sensible heat flux simulation, the Mosaic LSM is the most skillful, followed by Noah, MME, and VIC, respectively. Lastly, for the latent heat flux simulation, MME is the most skillful, followed

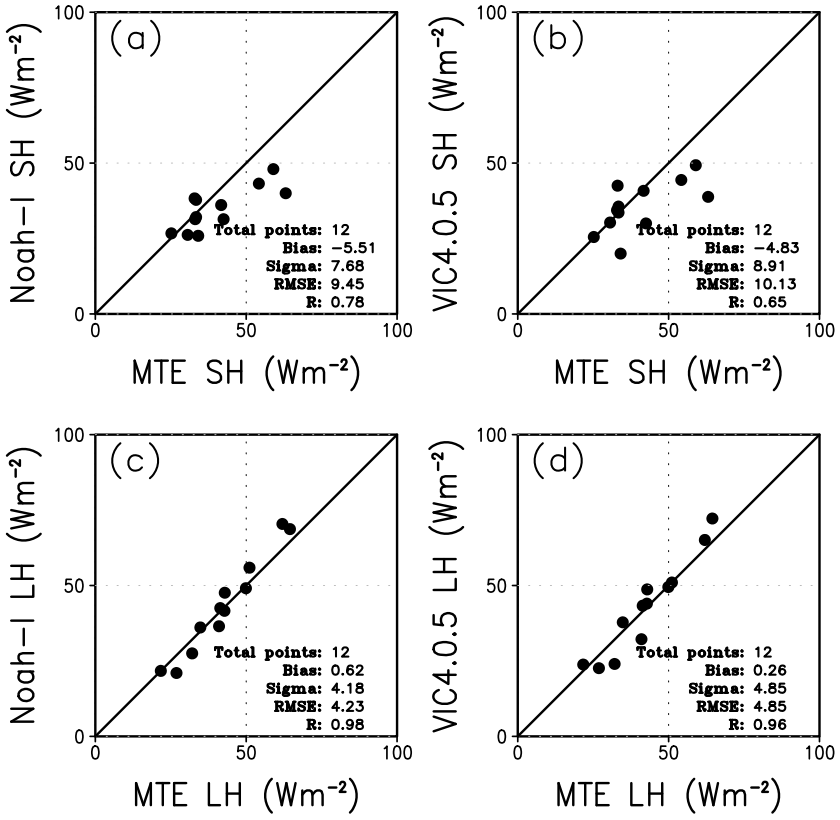

Figure 14. Scatterplots and their statistical metrics for (a and b) mean annual sensible heat flux (SH) and (c and d) latent heat flux (LH) of MTE reference versus NLDAS-2 simulations by the experimental Noah-I (Figures 14a and 14c) and experimental VIC 4.0.5 (Figures 14b and 14d).

by Noah, VIC, and Mosaic, respectively. Additionally in Table 4, all three LSMs and their ensemble mean have notably lower skills over the wet LMRFC and SERFC basins when compared to the other RFCs. These two RFC regions are considered to be water-unlimited and energy-limited regions in general (except during very substantial drought episodes).

Obviously and surprisingly in Table 4, all three LSMs and their ensemble mean show low AC values for R_n, although this varies somewhat by RFC and model. The primary reason is that all the LSMs have low AC for USWR at all 12 RFCs, in particular at LMRFC and SERFC (Table 5). In contrast to USWR, all the LSMs show relatively large AC values in Table 5 for the remaining three radiation fluxes (DSWR, ULWR, and DLWR). Interestingly, despite the small AC values for simulated USWR, quite large correlation values are apparent in Table 6 for USWR at all 12 RFCs, even at LMRFC and SERFC. Hence, the simulated USWR captures the seasonal-to-annual cycle of USWR reasonably well, but not the month-to-month variability.

3.2.3. Analysis of Intrabasin Differences in Monthly Anomaly Correlation and Taylor Skill Score

The analysis in sections 3.2.1 and 3.2.2 shows that the statistical metrics and model performances are basin dependent. For example, the LSMs perform better in LH simulations over dry basins (e.g., CBRFC and WGRFC) and better in SH simulations over wet basins (e.g., OHRFC and SERFC). However, it remains to be shown in this section how well the models perform simulating SH and LH at different areas within a given basin. In this section, an intrabasin difference analysis focusing on the Taylor skill score (*S*) and AC is used to investigate the models' performance within basins.

In Figure 13, clear differences are evident in the spatial distribution of *S* and AC within each basin. For example, in the WGRFC there are larger *S* values of SH over the western portion than over the eastern portion for all three LSMs. In the CNRFC there are smaller AC values of LH over the western portion than over the eastern portion for all three LSMs. Thus, as one would anticipate, significant intrabasin differences exist within all

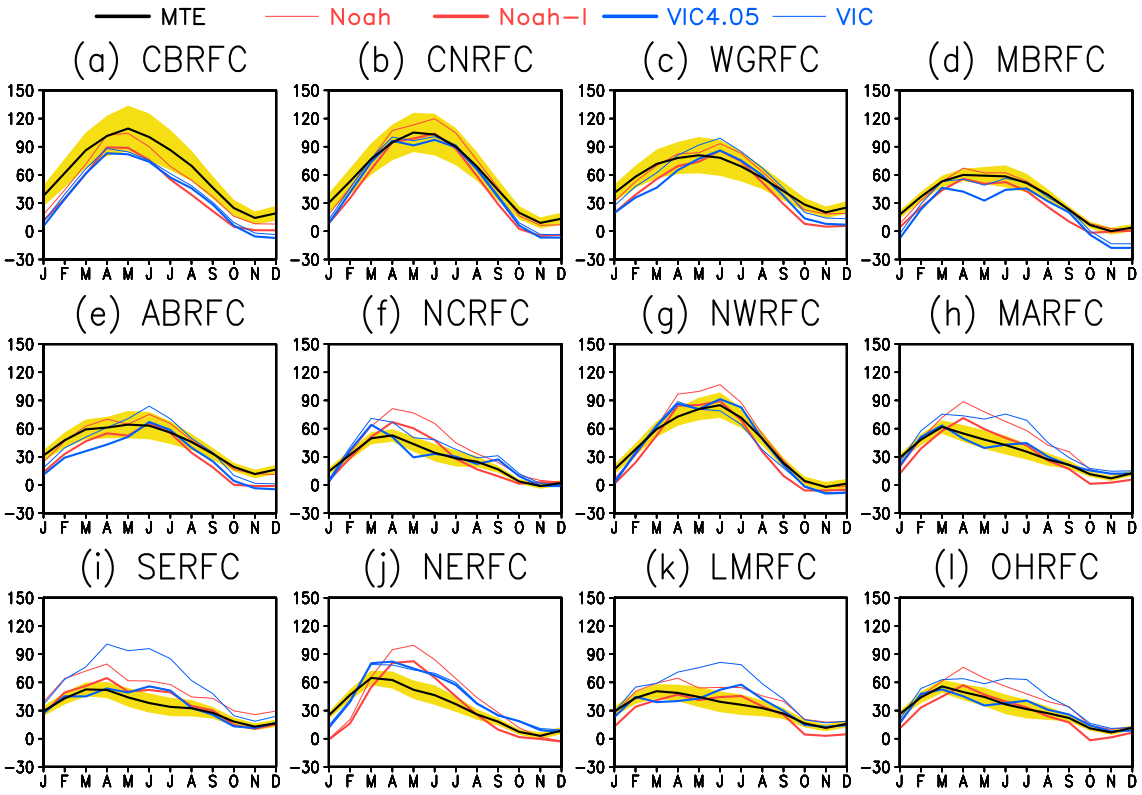

Figure 15. Mean annual cycle of monthly mean sensible heat flux (SH) (unit: W m^{-2}) from MTE reference (black) and NLDAS-2 simulated by the operational Noah (thin red line), operational VIC (thin blue line), experimental Noah-I (thick red line), and experimental VIC 4.0.5 (thick blue line).

the RFCs for all the LSMs and variables examined, although the spatial variability differs from basin to basin, model to model, and variable to variable.

The results also show that AC is significant for both SH and LH over most regions over the CONUS, although its value is low in some regions. For the SH simulations, there is no significant AC over the west and the northeast for VIC, and there is also no significant AC over the northeast and mountainous region for Noah. For the LH simulations, both Mosaic and VIC have insignificant AC values over the southeast. Those regions and models with insignificant AC values need to be further investigated in the future from the viewpoint of forcing data accuracy, model physics and model parameters, as well as the accuracy and reliability of the LH and SH anomalies in the validating data sets.

3.3. Evaluation of Research NLDAS-2 Energy Budget Components

The operational NLDAS-2 is producing real-time products to support U.S. operational drought monitoring and prediction activities. To advance and upgrade the next phase of the operational NLDAS-2 system, a research NLDAS-2 system is being tested at NCEP/EMC, from which surface energy components were generated from 1 January 1979 to 31 December 2007. Although this research version may not be transitioned exactly in its present configuration to the next phase of the operational NLDAS, the assessment of its surface energy components is helpful to improve the LSMs that will be used in the next operational phase.

An examination of mean annual SH and LH clearly shows large improvements in the research versions (e.g., Noah-I and VIC4.0.5, Figure 14) when compared to the corresponding operational versions (Noah and VIC, Figure 6). The correlation coefficient for SH is increased from 0.29 to 0.65 in the VIC4.0.5 simulation. The LH bias is reduced from -5.5 W m^{-2} to 0.6 W m^{-2} in the Noah-I simulation and from -0.48 W m^{-2} to 0.30 W m^{-2} in VIC4.0.5 simulation. The SH simulated by Noah-I and VIC4.0.5 is closer to the MTE reference values at all 12 RFCs when compared to the operational Noah and VIC, in particular for the warm season and at OHRFC, MARFC, LMRFC, and SERFC (Figure 15). Moreover, in Figure 16 it is evident that the Noah-I and VIC4.0.5-simulated LH values have increased and become closer to the MTE reference values (although VIC4.0.5 shows some positive biases in summer at NCRFC and

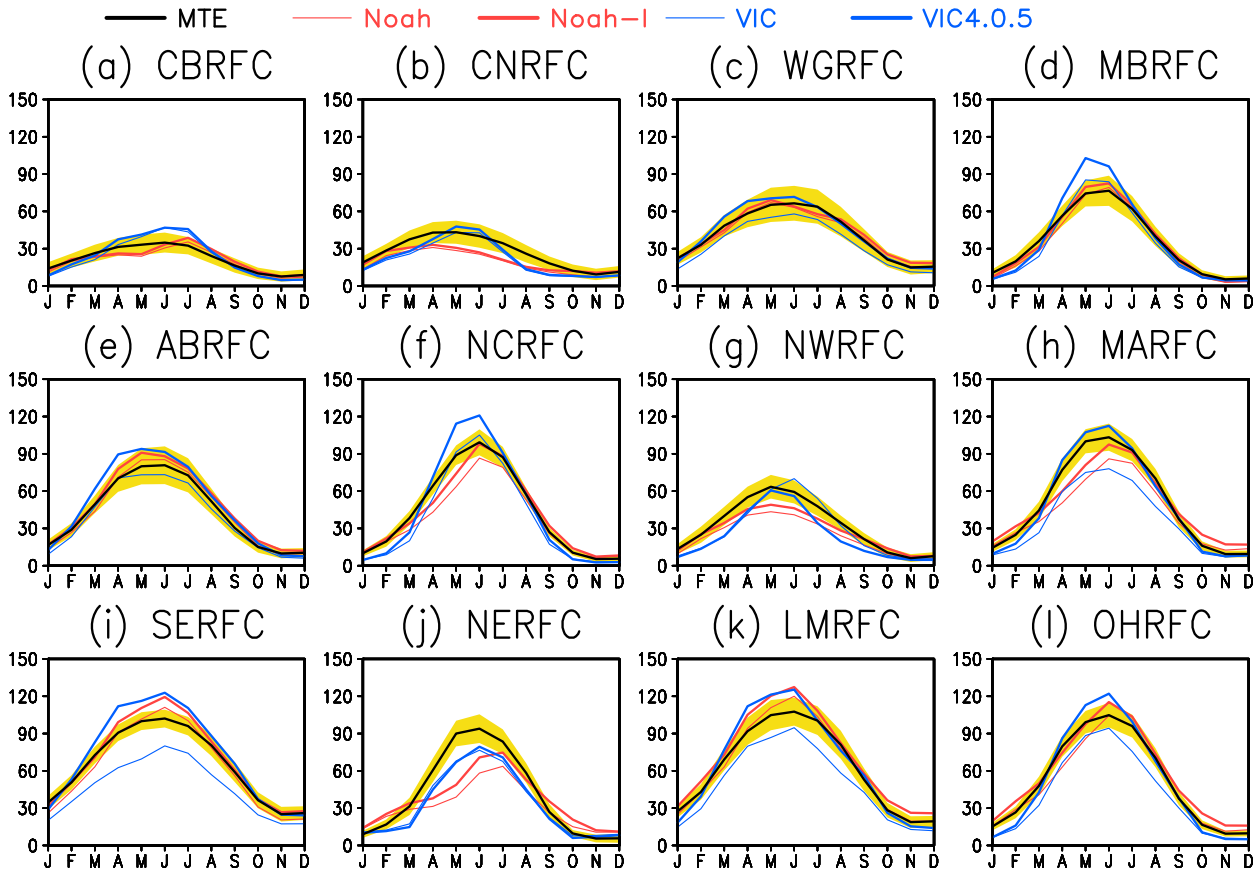

Figure 16. Mean annual cycle of monthly mean latent heat flux (LH) (unit: W m^{-2}) for MTE reference (black) and NLDAS-2 simulated by the operational Noah (thin red line), operational VIC (thin blue line), experimental Noah-I (thick red line), and experimental VIC 4.0.5 (thick blue line).

OHRFC). These improvements are further confirmed by comparison of the Taylor skill score for the simulation of SH (Table 7). For the Noah LSM, the Taylor skill score is increased at seven of 12 RFCs, and for the VIC LSM, it is increased at eight of 12 RFCs, suggesting that the SH simulations are improved at most RFCs with the research model versions. These improvements mainly come from either model physics upgrade (i.e., Noah-I) or model soil and hydrological parameters tuning procedure (i.e., VIC4.0.5). Therefore, not surprisingly, model physics improvements and refinements of the soil and hydrological parameters are two pathways to attain improved model simulations for the next operational era of NLDAS.

Table 7. The Taylor Skill Score Calculated From Monthly MTE Reference and NLDAS-2-Simulated SH (From Both Operational and Research NLDAS-2 Systems) for the 12 CONUS RFCs Over the 24.5 year Period (July 1983 to December 2007)[a]

RFC Label	Noah	Noah-I	VIC	VIC4.0.5
CBRFC	**0.99**	**0.99**	**0.99**	**0.99**
CNRFC	0.95	**0.97**	**0.97**	**0.97**
WGRFC	**0.92**	0.89	0.86	0.88
MBRFC	0.98	**0.99**	0.94	0.95
ABRFC	**0.94**	0.92	0.84	0.89
NCRFC	0.79	0.89	0.90	**0.93**
NWRFC	0.92	0.96	**0.97**	0.95
MARFC	0.79	0.87	0.81	**0.95**
SERFC	**0.87**	0.84	0.48	0.82
NERFC	0.66	0.77	**0.89**	0.88
LMRFC	**0.88**	0.85	0.62	0.83
OHRFC	0.84	0.91	0.83	**0.94**

[a]The bold font denotes the maximum value for each RFC.

4. Summary and Conclusion

The main components of the land surface energy budget calculated from the operational and research NLDAS-2 systems are analyzed at 12 RFCs. Model-computed surface energy estimates (i.e., sensible and latent heat flux) are compared with

the estimates from the MTE method [*Jung et al.*, 2009]. The satellite estimates from the GEWEX/SRB project [*Zhang et al.*, 2013, 2014] are used for the evaluation of the downward and upward shortwave and longwave radiation, as well as the net radiation. The analysis offers a description of the basic physical aspects of the surface energy budget as viewed from both models and independent estimates. The discussion is focused on both model successes as well as model problems that still need to be addressed. The results are summarized as follows:

1. NLDAS-2 radiative forcing fields (DSWR and DLWR) have high anomaly correlation skills (>0.78) and small errors (~10 W m^{-2}) with respect to the GEWEX/SRB reference data sets at all 12 RFCs, although NLDAS-2 DSWR is somewhat too large, particularly in the summer.
2. All three LSMs in the operational NLDAS-2 system simulate ULWR quite well, with high anomaly correlations and small errors with respect to the GEWEX/SRB. In contrast, all models show lower anomaly correlations and moderately positive biases of USWR when compared with GEWEX/SRB. The low anomaly correlation (AC) of the simulated USWR is due to overly large intermonth variability in the simulated surface albedo in all three LSMs, especially in the cool season in the northern RFCs typically characterized by persistent snow cover. The high bias of simulated USWR arises from the high bias in both (a) the simulated surface albedo and (b) the NLDAS-2 DSWR input forcing.
3. Owing to the low AC of the simulated USWR cited above, the R_n simulated by all three LSMs in the operational NLDAS-2 also shows low anomaly correlation. However, the LSMs reasonably simulate the magnitude of R_n with relatively small bias at all the RFCs, except for CNRFC and CBRFC where all three LSMs feature a low R_n bias.
4. All three LSMs in the operational NLDAS-2 reasonably simulate the mean annual cycle and monthly variability of SH and LH at all 12 RFCs with respect to the validating MTE values. However, there are notable instances of smaller anomaly correlations of LH, especially over the moist basins, and to a lesser degree over the semiarid basins, than over the moisture transition basins of the central CONUS (Great Plains). Overall, results show that Mosaic has a high LH bias compared to MTE at all 12 RFCs. A low LH bias is typical in the VIC and Noah LSMs.
5. Two models (i.e., Noah-I and VIC4.0.5) in the research NLDAS-2 system moderately improve their SH and LH simulations with respect to the MTE reference data set for both mean annual values and monthly mean annual cycles when compared to their operational counterparts. This is due to updates in some model parameters [*Xia et al.*, 2014c; *Troy et al.*, 2008].

Notwithstanding the limited improvements in Noah-I and VIC4.0.5, this study suggests that room remains to enhance the accuracy of the surface energy budget components in the NCEP operational NLDAS-2 system. They include the following: (1) improving NLDAS-2 forcing data such as downward shortwave and longwave radiation, 2 m air temperature and specific humidity, and 10 m wind speed; (2) further improving model physical processes and model parameters; and (3) deriving improved fields of vegetation fraction and surface albedo incorporating improved weekly, monthly, and seasonal variability, including especially that of snow cover effects. NASA/GSFC, NCEP/EMC, and the National Water Center are also collaborating on adding new and upgraded LSMs (including replacement of the NASA Mosaic LSM with the NASA Catchment LSM and upgrading SAC to SAC-HTET), as well as assimilating satellite remote sensing data of near-surface and deeper water states into the operational NLDAS system by means of the NASA-developed Land Information System (LIS, http://lis.gsfc.nasa.gov/) [*Kumar et al.*, 2006; *Peters-Lidard et al.*, 2007].

This study shows that large difference between simulated and SRB upward solar radiation is most likely due to large difference between SRB and model-integrated surface albedo. The major reason in the warm season is that the LSMs appear to assign larger bare soil and vegetation albedo values than inferred from the satellite-based estimates. The major reason in the cold season for the four most northern RFCs (having seasonally persistent snow cover) is that all three LSMs overamplify the effects of snow cover on the surface albedo (as compared to values inferred from the satellite-based estimates), likely due to the assignment of overly large values of both (1) maximum albedo for deep snow and (2) snow cover fraction with increasing snow depth. Since model-integrated albedo is a function of soil texture, vegetation type, bare soil and vegetation fraction, bare soil and vegetation albedo, snow cover fraction, and maximum snow albedo for deep snow, a comprehensive comparison of these parameters among models and against in situ observations and satellite estimates needs to be done in the future.

Additional upgrades to all the NLDAS models are now available. Noah 2.8 and Noah-I can be replaced with Noah 3.6 and Noah-MP [*Niu et al.*, 2011; *Yang et al.*, 2011], which add multilayer snowpack, vegetation

dynamics, ground water, and irrigation modules; VIC4.03 and VIC4.05 can be upgraded to VIC4.1.2, which updates model physical processes and requires additional tuning of model parameters (see http://www.hydro.washington.edu/Lettenmaier/Models/VIC/Overview/ModelOverview.shtml); Mosaic can be replaced by the NASA Catchment model, which updates model physics [*Koster et al*., 2000]; and SAC and SAC-Clim can be replaced by SAC-HTET, which adds soil heat transfer and an energy cycle simulation [*Koren et al*., 2007, 2010, 2014].

Furthermore, these upgraded models have been incorporated into the Land Information System framework (LIS) [*Kumar et al*., 2006; *Peters-Lidard et al*., 2007], which also supports data assimilation (e.g., soil moisture and snow data) to further improve the simulation of land surface energy components. A preliminary test has demonstrated significant improvements for LH simulation and other state variables as well [*Kumar et al*., 2014]. These efforts described above are being included in the next generation NLDAS development. We believe that the upgraded models with actual data assimilation can more accurately simulate land surface energy components. The evaluation will be executed by using the same framework as used here, and the results will be reported in the near future.

Acknowledgments

We thank Martin Jung of the Maxi Planck Institute for providing the MTE SH and LH data and the NASA Langley Research Center for providing the SRB satellite radiation observations. We also thank Yu-Tai Hou at NCEP/EMC, Jiafu Mao at Oak Ridge National Laboratory, and two anonymous reviewers whose comments greatly improved the quality of this manuscript. This work was supported by the NOAA Climate Program Office MAPP program. NLDAS-2 products can be freely obtained from both the NCEP/EMC ftp site (ftp://ldas.ncep.noaa.gov/nldas2/retrospective/) and the NASA NLDAS website (http://ldas.gsfc.nasa.gov/nldas/NLDASnews.php). The NASA GEWEX/SRB radiation products can be obtained from the NASA website (http://gewex-srb.larc.nasa.gov/). The gridded FLUXNET sensible and latent heat flux products, including their uncertainty estimates, can be obtained (via Martin Jung) from https://www.bgc-jena.mpg.de/geodb/projects/Home.php. These websites serve as valuable resources for readers and data users who seek to leverage these data sets to perform research investigations.

References

Baldocchi, D. D., et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem–scale carbon dioxide, water vapor, and energy flux densities, *Bull. Am. Meteorol. Soc.*, *82*, 2415–2434.

Bentsen, M., et al. (2013), The Norwegian Earth System Model, NorESM1-M—Part 1: Description and basic evaluation of the physical climate, *Geosci. Model Dev.*, *6*, 687–720.

Berbery, E. H., K. E. Mitchell, S. Benjamin, T. Smirnova, H. Ritchie, R. Hogue, and E. Radeva (1999), Assessment of land-surface energy budgets from regional and global models, *J. Geophys. Res.*, *104*, 19,329–19,348.

Berg, A. A., J. S. Famiglietti, J. P. Walker, and P. R. Houser (2003), Impact of bias correction to reanalysis products on simulations of North American soil moisture and hydrological fluxes, *J. Geophys. Res.*, *108*(D16), 4490, doi:10.1029/2002JD003334.

Betts, A. K., P. Viterbo, and E. Wood (1998), Surface energy and water balance for the Arkansas-Red basin from the ECMWF reanalysis, *J. Clim.*, *11*, 2881–2897.

Betts, A. K., J. H. Ball, and P. Viterbo (1999), Basin-scale surface water and energy budgets for the Mississippi from the ECMWF reanalysis, *J. Geophys. Res.*, *104*, 19,293–19,306.

Boone, A. A., I. Poccard-Laclercq, Y. Xue, J. Feng, and P. de Rosnay (2009), Evaluation of the WAMME model surface fluxes using results from the AMMA land-surface model intercomparison project, *Clim. Dyn.*, *35*, 127–142, doi:10.1007/s00382-009-0653-1.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system: Conceptual models for digital computer, *Tech. Rep*., Joint Fed.-State River Forecast Cent., U.S. Natl. Weather Serv. And Calif. Dep. of Water Resour., Sacramento, Calif.

Cai, X., Z.-L. Yang, Y. Xia, M. Huang, H. Wei, L. R. Leung, and M. B. Ek (2014), Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed, *J. Geophys. Res. Atmos.*, *119*, 13,751–13,770, doi:10.1002/2014JD022113.

Cosgrove, B. A., et al. (2003), Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8842, doi:10.1029/2002JD003118.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.

Deardorff, J. W. (1978), Efficient prediction of ground surface temperature and moisture inclusion of a layer of vegetation, *J. Geophys. Res.*, *83*, 1889–1903.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rodgers, P. Grunman, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, *108*(D22), 8851, doi:10.1029/2002JD003296.

Feng, X., and P. Houser (2008), An investigation of GSWP-2 Mississippi River basin surface water and energy budgets, *J. Geophys. Res.*, *113*, D15118, doi:10.1029/2007JD009144.

Getirana, A. C. V., et al. (2014), Water balance in the amazon basin from a land surface model ensemble, *J. Hydrometeorol.*, *15*, 2586–2614.

Hinkelman, L. M., P. W. Stackhouse, B. A. Wielicki, T. Zhang, and S. R. Wilson (2009), Surface insolation trends from satellite and ground measurements: Comparisons and challenges, *J. Geophys. Res.*, *114*, D00D20, doi:10.1029/2008JD011004.

Jarvis, P. G. (1976), The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field, *Philos. Trans. R. Soc. London Ser. B.*, *273*, 593–610.

Joyce, R., J. Janowiak, P. Arkin, and P. Xie (2004), CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution, *J. Hydrometereol.*, *5*, 487–503.

Jung, M., M. Reichstein, and A. Bondeau (2009), Towards global empirical upscaling of FLUXNET eddy covariance observations: Validation of a model tree ensemble approach using a biosphere model, *Biogeosciences*, *6*, 2001–2013, doi:10.5194/bg-6-2001-2009.

Jung, M., et al. (2010), Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, *467*, 951–954.

Jung, M., et al. (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *J. Geophys. Res.*, *116*, G00J07, doi:10.1029/2010JG00156.

Koren, V., M. Smith, Z. Cui, and B. Cosgrove (2007), Physically-based modifications to the Sacramento Soil Moisture Accounting Model: Modeling the effects of frozen ground on the rainfall-runoff process, *NOAA Tech. Rep., No.52*, 47 pp., National Weather Service, Silver Spring, Maryland. [Available at http://www.nws.noaa.gov/oh/hrl/hsmb/hydrology/PBE_SAC-SMA/index.html.]

Koren, V., M. Smith, Z. Cui, B. Cosgrove, K. Werner, and R. Zamora (2010), Modification of Sacramento Soil Moisture Accounting Heat Transfer Component (SAC-HT) for enhanced evapotranspiration, *NOAA Tech. Rep., No.53*, 72 pp., National Weather Service, Silver Spring, Md. [Available at http://www.nws.noaa.gov/oh/hrl/hsmb/hydrology/PBE_SAC-SMA/index.html.]

Koren, V., M. Smith, and Z. Cui (2014), Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process, *J. Hydrol.*, *519*, 3475–3491.

Koster, R., and M. Suarez (1994), The components of a SVAT scheme and their effects on a GCM's hydrological cycle, *Adv. Water Resour.*, *17*, 61–78.
Koster, R., and M. Suarez (1996), Energy and water balance calculations in the Mosaic LSM, *NASA Tech. Memo.*, 104606, vol. 9, 60 pp.
Koster, R. D., and M. J. Suarez (1992), Modeling the land surface boundary in climate models as a composite of independent vegetation stands, *J. Geophys. Res.*, *97*, 2697–2715, doi:10.1029/91JD01696.
Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a GCM: Part 1. Model structure, *J. Geophys. Res.*, *105*, 24,809–24,822.
Kumar, S. V., et al. (2006), Land Information System: An interoperable framework for high resolution land surface modeling, *Environ. Modell. Software*, *21*, 1402–1415.
Kumar, S. V., et al. (2014), Assimilation of remotely sensed soil moisture and snow depth retrievals for drought estimation, *J. Hydrometeorol.*, *15*, 2446–2469.
Liang, S., K. Wang, X. Zhang, and M. Wild (2007), Review on estimation of land surface radiation and energy budgets from ground measurement, remote sensing and model simulations, *IEEE J. Selec. Top. Appl. Earth Obs. Remote Sens.*, *3*, 225–239.
Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for GCMs, *J. Geophys. Res.*, *99*, 14,415–14,428.
Lin, Y., and K. Mitchell (2005), The NCEP stage II/IV hourly precipitation analyses: Development and applications, preprint paper 1.2 presented a19th Conference on Hydrology, Am. Meteorol. Soc., San Diego, Calif., 9–13 Jan.
Livneh, B., Y. Xia, K. E. Mitchell, M. B. Ek, and D. P. Lettenmaier (2010), Noah LSM snow model diagnostics and enhancements, *J. Hydrometeorol.*, *11*, 721–738.
Mahrt, L., and M. Ek (1984), The influence of atmospheric stability on potential evaporation, *J. Climate Appl. Meteorol.*, *23*, 222–234.
Meng, J., R. Yang, H. Wei, M. Ek, G. Gayno, P. Xie, and K. Mitchell (2012), The land surface analysis in the NCEP Climate Forecast System Reanalysis, *J. Hydrometeorol.*, *13*, 1621–1630.
Mesinger, F., et al. (2006), North American Regional Reanalysis, *Bull. Am. Meteorol. Soc.*, *87*, 343–360.
Mitchell, K. E., et al. (1999), The GCIP Land Data Assimilation System (LDAS) project: Now underway, *GEWEX News*, *9*(1), 7–11.
Mitchell, K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res.*, *109*, D07S90, doi:10.1029/2003JD003823.
Monteith, J. L. (1965), Evaporation and environment, *Symp. Soc. Exp. Biol.*, *19*, 205–224.
Niu, G.-Y., et al. (2011), The community Noah land surface model with multi-physics options. Part 1: Model descriptions and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*, D12109, doi:10.1029/2010JD015139.
Ohmura, A., et al. (1998), Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research, *Bull. Am. Meteorol. Soc.*, *79*, 2115–2136.
Peters-Lidard, C. D., et al. (2007), High-performance Earth system modeling with NASA/GSFC's Land Information System, *Innovations Syst. Software Eng.*, *3*, 157–165.
Peters-Lidard, C. D., S. V. Kumar, D. Mocko, and Y. Tian (2011), Estimating evapotranspiration with land data assimilation systems, *Hydrol. Processes*, *25*, 3979–3992, doi:10.1002/hyp.8387.
Pinker, R. T., et al. (2003), Surface radiation budgets in support of the GEWEX Continental-Scale International Project (GCIP) and the GEWEX Americas Prediction Project (GAPP), including the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, *108*(D22), 8844, doi:10.1029/2002JD003301.
Raschke, E., S. Kinne, and P. W. Stackhouse (2012), GEWEX Radiative Flux Assessment (RFA) Volume 1: Assessment, WCRP Rep. 19/2012, 273 pp., World Climate Research Programme (WCRP), Geneva, Switzerland. [Available at http://www.wcrp-climate.org/documents/GEWEX%20RFA-Volume%201-report.pdf.]
Roads, J., et al. (2003), GCIP water and energy budget synthesis (WEBS), *J. Geophys. Res.*, *108*, D8609, doi:10.1029/2002JD002583.
Robock, A., et al. (2003), Evaluation of the North American Land Data Assimilation System (NLDAS) over the Southern Great Plains during the warm season, *J. Geophys. Res.*, *108*(D22), 8846, doi:10.1029/2002JD003245.
Saha, S., et al. (2014), The NCEP Climate Forecast System version 2, *J. Clim.*, *27*, 2185–2208.
Schaake, J. C., et al. (2004), An intercomparison of soil moisture fields in the North American Land Data Assimilation System (NLDAS), *J. Geophys. Res.*, *109*, D01S90, doi:10.1029/2002JD003309.
Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986), A Simple Biosphere Model (SIB) for use within General Circulation Models, *J. Atmos. Sci.*, *43*, 505–531.
Sheffield, J., et al. (2003), Snow process modeling in the North American Land Data Assimilation System (NLDAS): 1. Evaluation of model-simulated snow cover extent, *J. Geophys. Res.*, *108*(D22), 8849, doi:10.1029/2002JD003274.
Stegehuis, A. I., A. J. Teuling, P. Ciais, R. Vautard, and M. Jung (2013), Future European temperature change uncertainties reduced by using land heat flux observations, *Geophys. Res. Lett.*, *40*, 2242–2245, doi:10.1002/grl.50404.
Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*, 7183–7192.
Troy, T. J., and E. F. Wood (2009), Comparison and evaluation of gridded radiation products across northern Eurasia, *Environ. Res. Lett.*, *4*, 1–7.
Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resour. Res.*, *44*, W09411, doi:10.1029/2007WR006513.
Velpuri, N. M., G. B. Senay, R. K. Singh, S. Bohms, and J. P. Verdin (2013), A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: Using point and gridded FLUXNET and water balance ET, *Remote Sens. Environ.*, *139*, 35–49.
Wei, H., Y. Xia, K. E. Mitchell, and M. B. Ek (2013), Improvement of the Noah land surface model for warm season processes: Evaluation of water and energy flux simulation, *Hydrol. Processes*, *27*, 297–303.
Wild, M., J. Grieser, and C. Schär (2008), Combined surface solar brightening and greenhouse effect support recent intensification of the global land-based hydrological cycle, *Geophys. Res. Lett.*, *35*, L17706, doi:10.1029/2008GL034842.
Wood, E. F., D. P. Lettenmaier, X. Liang, B. Nijssen, and S. W. Wetzel (1997), Hydrological modeling of continental-scale basins, *Annu. Rev. Earth Planet Sci.*, *25*, 279–300.
Xia, Y., et al. (2012), Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*, D03109, doi:10.1029/2011JD016048.
Xia, Y., et al. (2013), Validation of Noah-simulated soil temperature in the North American Land Data Assimilation System Phase 2, *J. Appl. Meteorol. Climatol.*, *52*, 455–471.
Xia, Y., J. Sheffield, M. B. Ek, J. Dong, N. Chaney, H. Wei, J. Meng, and E. F. Wood (2014a), Evaluation of multi-model simulated soil moisture in NLDAS-2, *J. Hydrol.*, *512*, 107–125.
Xia, Y., M. T. Hobbins, Q. Mu, and M. B. Ek (2014b), Evaluation of NLDAS-2 evapotranspiration against tower flux site observations, *Hydrol. Processes*, doi:10.1002/hyp.10299.

Xia, Y., C. D. Peter-Lidard, M. Huang, H. Wei, and M. Ek (2014c), Improved NLDAS-2 Noah-simulated hydrometeorological products with an interim run, *Hydrol. Processes*, *29*, 780–792, doi:10.1002/hyp.10190.

Xia, Y., et al. (2015a), Basin-Scale Assessment of the land surface water budget components in NCEP operational and research NLDAS-2 system, *J. Geophys. Res. Atmos.*, *120*, doi:10.1029/2015JD023733.

Xia, Y., M. Ek, Y. Wu, T. Ford, and S. Quiring (2015b), Comparison of NLDAS-2 simulated and NASMD observed daily soil moisture. Part I: Comparison and analysis, *J. Hydrometeorol.*, *16*, 1962–1980.

Yang, R., M. B. Ek, and J. Meng (2015), Surface water and energy budgets for the Mississippi River Basin in three NCEP reanalyses, *J. Hydrometeorol.*, doi:10.1175/JHM-D-14-0056.1.

Yang, Z.-L., et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins, *J. Geophys. Res.*, *116*, D12110, doi:10.1029/2010JD015140.

Zhang, T., P. W. Stackhouse, S. K. Gupta, S. J. Cox, M. Colleen Mikovitz, and L. M. Hinkelman (2013), The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach, *J. Quant. Spectrosc. Radiat. Transfer*, *122*, 127–140.

Zhang, T., P. W. Stackhouse, J. S. Gupta, S. J. Cox, and J. Colleen Mikovitz (2014), The validation of the GEWEX SRB surface longwave flux data products using BSRN measurements, *J. Quant. Spectrosc. Radiat. Transfer*, doi:10.1016/j.jqsrt.2014.07.013.

Zheng, W., H. Wei, Z. Wang, X. Zeng, J. Meng, M. Ek, K. Mitchell, and J. Derber (2012), Improvement of daytime land surface skin temperature over arid regions in the NCEP GFS model and its impact on satellite data assimilation, *J. Geophys. Res.*, *117*, D06117, doi:10.1029/2011JD015901.